Stochastic Programming
for Home Energy Management System Optimization

Master's Thesis

Natalia Escobosa Pineda

Escola Tècnica Superior d'Enginyeria Industrial de Barcelona

Universitat Politècnica de Catalunya

Supervisors: Andreas Sumper
Pol Olivella-Rosell

Barcelona, September 2018

Abstract

The European Union set the 2020 goals to reduce GHG emissions, increase energy efficiency and decarbonize the energy supply. Home Energy Management Systems (HEMS) are a source of flexibility and a cost-effective strategy to pursue the integration of Renewable Energy Sources (RES) However, the behavior of the components involved in the residential energy consumption is highly uncertain which poses a challenge for HEMS.

A strategy to tackle such uncertain parameters is stochastic optimization. Even though its formulation dates back decades, only the recent rise of adequate technology and software have made possible its implementation. This work proposes to use stochastic optimization in HEMS.

The requirements for Stochastic Programming (SP) consist of a mathematical model, a scenario tree and different data instances. The mathematical model is based on the INVADE project and was implemented using Pyomo. The stochastic formulation of the problem was done with PySP, the Pyomo extension for SP. The scenario tree was created using NetworkX and the data instances were formulated with Dplyr.

A household with an inflexible load, PV generation, a battery and a connection to the grid was considered. Three case studies were analyzed, to gain insight on the impact of different stochastic parameters. Furthermore, a sensitivity analysis regarding the nature of the stochastic input data was performed. Afterwards, a detailed description and analysis regarding the use of the software was done. Finally, the environmental impact of the project was assessed.

In conclusion the stochastic and deterministic formulations are equivalent for the present work due to the high flexibility of the grid. The Value of Stochastic Solution (VSS) was of around -0.04 €. However, this value increases with increasing standard deviation (σ) of the input data.

The algorithm schedules the HEMS components in response to the market price signals. The main source of flexibility is the grid, followed by the battery. Generation curtailment is also attractive and is scheduled in all simulations. Lastly, feed in to the grid is the least attractive flexibility mechanism. The expected flexibility was of 0.381 € per day.

R and Python proved to be simple and powerful. Furthermore, Pyomo is ideal to translate models into python objects. PySP has its advantages and drawback, although some of the last ones were circumvent by using NetworkX and Pandas.

This algorithm has an estimated mitigation potential of 1gCO_{2eq}/per day for the case studies analyzed. On the other hand, the environmental impact created during the realization of this project was of 115 kgCO_{2eq}.

ETSEIB

Index

List of Figures

List of Tables

Glossary

Renewable Energy includes bioenergy, direct solar energy, geothermal energy, hydropower, ocean energy and wind energy as defined by the IPCC

Home Energy Management Systems (HEMS): "*demand response tools that shift and curtail demand to improve the energy consumption and production profile of a dwelling on behalf of a consumer. HEMS usually create optimal consumption, production [and/or storage] schedules by considering multiple objectives such as energy costs, environmental concerns, load profiles, and consumer comfort "* [1].

Pyomo: It stands for "Python Optimization Modelling Objects", and as its name states it, it is a python-based tool used for mathematical modelling [2].

PySP: An extension of the Pyomo package designed to model objects for SP

Expected Monetary Value: "*weighted average of the payoffs for a decision alternative*

Progressive Hedging (PH): Formulation for the stochastic models of the problem proposed by Rockafellar and Wets in 1991. It is classified as a horizontal decomposition strategy [3].

Extensive Form (EF): Formulation for the model of a stochastic problem. It explicitly describes each stage-decision variable for all scenarios. It is considered the deterministic equivalent of a stochastic problem [3].

Standard Deviation (σ): "*a measure of the dispersion of a frequency distribution that is the square root of the arithmetic mean of the squares of the deviation of each of the class frequencies from the arithmetic mean of the frequency distribution"* [4] *.*

Variance: "*the squared of the standard deviation"* [5] *.*

1. Preface

The European Union has set targets to reduce their Greenhouse Gases (GHG) emissions by at least 20%, increase the share of renewable energy by 20% and save energy by 20% all by the year 2020. By setting these goals, the EU wants to fight climate change, air pollution, increase their energy security and reduce costs associated to energy [6].

1.1. Motivation

According to the Intergovernmental Panel on Climate Change (IPCC), human influence in the earth climate dynamics is clear. Since the 1950s, the effects of this influence have been observed: humanity has felt changes which are unprecedented over decades or even millennia in some cases. All of which pose a threat to human and natural systems due to the diverse and potential harming impacts [7]. Temperature has raised since the industrial area and great efforts have been done to settle which are the contributors to this rise in temperature and consequently influence in climate. As it can be observed in Figure 1, the main contributor that explains the change in temperature is due to the anthropogenic emissions GHG [7]. These have been the highest in recorded history for the past decades.

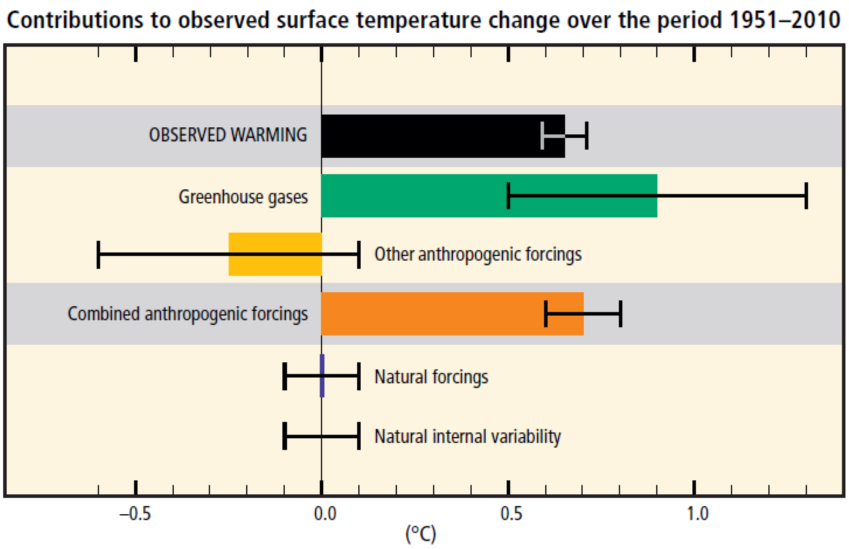

Figure 1. Assessed likely warming trends over the 1951-2010 period. Anthropogenic and natural forcings [7]

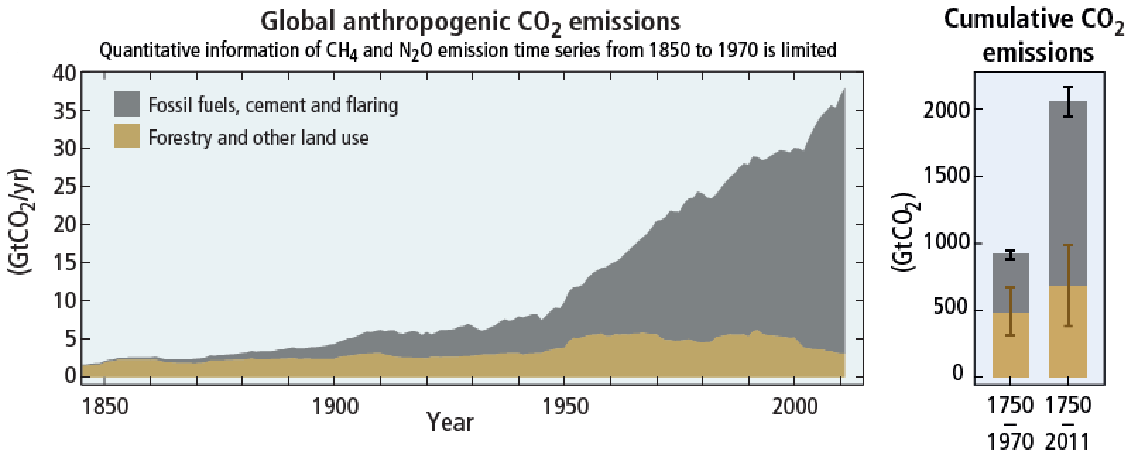

Figure 2. Main sources of global anthropogenic CO_2 emissions

Moreover, the IPCC recorded the origin of anthropogenic carbon dioxide and other GHG emissions. As it can be seen in Figure 2,since the 1950s fossil fuels have become the dominant source of anthropogenic emissions [7]. The energy sector accounts for around 35% of the GHG emissions [8].Therefore, huge efforts have been made to tackle and mitigate them in the energy sector. Decarbonization of the energy system is essential to attenuate the harmful consequences of climate change [8].

To assess the emission mitigation potential in the energy sector, an overall analysis to the energy system is needed. The energy system consists of all the processes that involve transforming a primary energy source and delivering it to an end user. In general, there are five steps of processes: Supply, Conversion, Transmission and/or Distribution, Demand and Storage [8]. The specific process is dependent on the energy source used and its final application.

As it is shown in Table 1Table 1. Distribution of energy carriers according to its source [9], oil products account for around 40% of the total final primary energy consumption, the main end user is transport. Electricity is the second most important carrier, followed closely by gas.

Table 1. Distribution of energy carriers according to its source [9]

Energy Source	Share of Energy Carriers in Total Final Consumption	Main End User
Coal and Peat	9.97%	Industry

Crude Oil	0.40%	Industry/Raw material
Oil products	41.30%	Transport
Gas	15.40%	Buildings and Industry
Geothermal, Solar, Wind, etc.	0.26%	Buildings
Waste and biofuels	12.87%	Buildings
Electricity	16.84%	Buildings and Industry
Heat	2.96%	Buildings

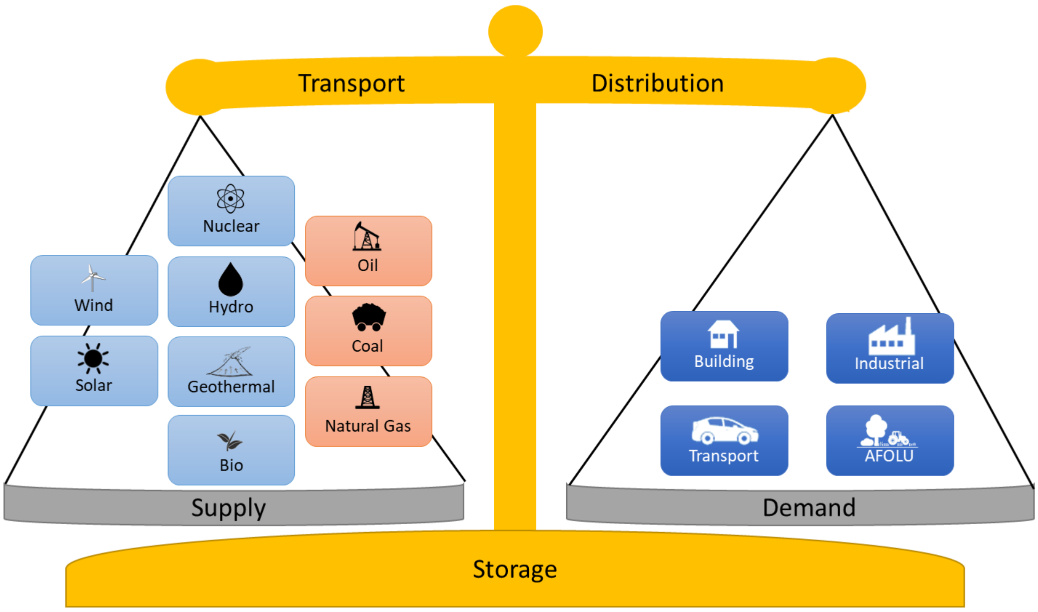

Figure 3. Representation of the balance of the power system

Figure 3 shows a representation of the electricity energy system or power system. Energy is harnessed from a diversity of sources to supply diverse needs. These sources include fossil fuels such as oil, coal and natural gas which have been the major dominant players in the last decades and have high GHG emissions. Further along the chain, energy is lost in all the transformation processes until is finally delivered to the different end users. The efficiency of the whole process for a fossil fuel power plant is of 37% [8]. As it is shown in Table 1, the main users are buildings and industry, however is expected that transport and Agriculture, Forestry and Other Land Use (AFOLU) will increase its share of electricity. Buildings are the main user of electricity with around 50% of the total consumption [10].

Some options proposed by the IPCC to reduce emissions in the electricity sector are:

- Improving energy efficiency in all the processes from energy conversion throughout the power system. This calls for better technologies while generating, better equipment to transport and convert the energy, less losses in storage technology as well as more efficient equipment in the end users.

- High emission fossil fuel switching such as the replacement of coal for oil or natural gas which have less emissions.

- Replacement of fossil fuels with low-GHG energy sources such as renewables or nuclear.

- Carbon Dioxide Capture and Storage.

In general, decarbonization of the energy supply is an objective shared by countries worldwide. Efforts have been made to achieve this goal, which have brought a range of different technologies to harness wind, sun, geothermal, waste, etc. All of which have benefits and drawbacks [8].

As it is shown in Figure 4, the share of RES is growing but is still small compared to the rest of the sources [11]. While analyzing Figure 5, specifically for electricity, the steady growth of RES is appreciated [12]. Decarbonization has happened faster in electricity generation than in the heat sector or liquid fuel sector [8]. Nonetheless, it is expected that the popularization of Electric Vehicles (EV) and the substitution of space heating by gas with space heating by electricity, the total energy mix will increase its electricity production by renewables and therefore decrease the consumption of other fossil fuels without a low-emission, low-cost substitute [8].

In consequence, huge efforts have made the cost of renewables drop, their performance improve and therefore a growth in the share of renewables for electricity generation. These resulted in over half of the new installed capacity worldwide is for RES. Furthermore, decentralization of the generation has increased for most of RES [8] as is appreciated in Figure 5. In 2012, Renewables accounted for 21% of the electricity mix, wind had a 5-fold increase and solar PV had a 25-fold increase from 2005 to 2012 [8].

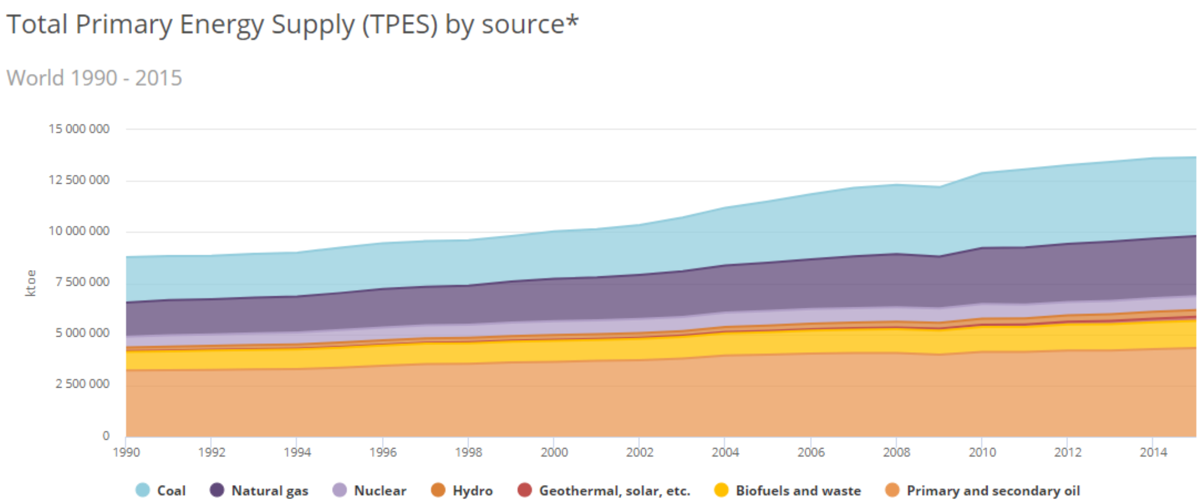

Figure 4. Evolution of the energy mix sources from 1990 to 2015 [11]

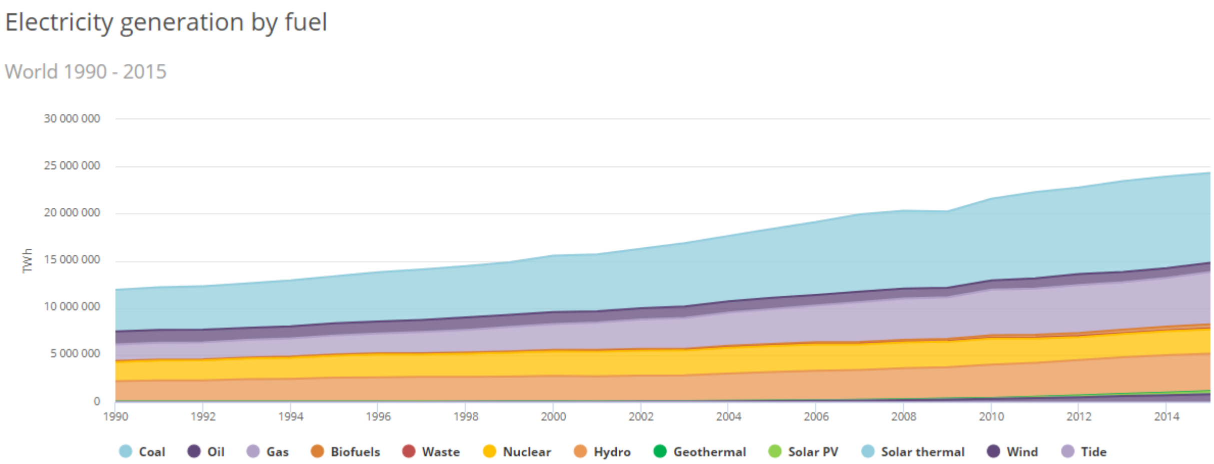

Figure 5. Evolution of the electricity mix sources from 1990 to 2015 [12]

There are other benefits to RES besides decarbonization of the energy supply such as reduction of air pollution, jobs creation, energy security and increase of energy access- the overall estimated potential of the different sources is at least 2.6 times the total primary energy demand of 2007 [8]. The drawbacks are being treated through public policies to increase their share in the electricity mix [8]. One of the most restricting characteristics of RES, such as wind, solar or hydro is that they are highly dependent on the location, time and technology. These restrictions are transmitted to the rest of the power system, reducing drastically the flexibility on the generation side, which creates a new challenge that needs to be addressed. Moreover, some technologies for harnessing RES potentials are mature enough to do a widespread depletion, but other are still in the R&D phase [8], one of them is to recuperate the flexibility of the system by using demand response.

1.2. Origin of the Project

Introduction of RES into the grid has brought up some problems and challenges: With less flexibility from the generation side, elevated costs due to RES technologies and distributed generation, there is a big interest in finding cost-efficient strategies for all these problems. An inexpensive strategy considered to get the required flexibility is using demand response from the consumer side. In order to cope with variable RES, distributed generation and lack of infrastructure, it is necessary to increase the flexibility, both in the supply and the load sides [13].

As part of the strategy to achieve the 2020 objectives, the European Union is presenting the INVADE project, whose objective is to study the different possibilities to increase the penetration and integration of RES into the power system and the electricity market. Specifically, there is a section of the INVADE project that focuses in the flexibility services offered by centralized, distributed and mobile storage to "*achieve optimal deployment of flexible energy storage in distribution systems*"[13].

At the center of the INVADE project is a cloud-based flexibility management system to integrate RES and EV, furthermore it is also included smart control of the domestic loads to help keep the reliability constraints during the day [14]. The purpose is to apply new technologies to existing infrastructure to benefit economically the end users and the system operators. Furthermore, the end users can have better services and the system operators can manage its resources optimally. The creation of new business models will aim to maximize profits for the different stakeholders involved as well as contributing to the 2020 targets [14].

The specific goals of the project are:

1) *"Design a flexibility management system using batteries that supports the distribution grid and electricity market while coping with grid limitations, uncertainty and variability with high penetration of renewable energy, electric vehicles and an increased number of diverse smart grid actors.*

2) *Develop a model for batteries including EVs focusing on prediction of batteries lifetime and impact factors contributing to life extension, and prepare a model for optimal sizing, positioning and scheduling of batteries in the distribution grid.*

3) *Deliver the Integrated INVADE Platform based on Flexibility Cloud enabling flexible management algorithms, functions and monitoring and control dashboards using Internet of Energy Things, Big data analytics and visualization techniques to provide real-time information and control tools to stakeholders applying data protection and cyber security principles by design.*

4) *Integrate the INVADE platform with existing infrastructure and systems in selected pilot sites in Bulgaria, Germany, Spain, Norway and the Netherlands and validate the platform through mobile, distributed, centralized and hybrid use cases in large-scale demonstrations in accordance with national and European regulations and standards.*

5) *Design innovative and competitive business models and verify them through planned activities such as analysis of users practices and behavior, deferral of grid investments, exploitation user group and dedicated workshops to enable monetary and social benefits for a full chain of stakeholders.*

6) *Engage with full chain stakeholders to support large scale deployment of INVADE within EEA and beyond and to build awareness of the project and its contribution to both climate change and energy efficiency targets.* "[14]

To comply with these objectives several work projects have been established, tackling different sections of the project. This thesis was developed within the objective of modelling the grid components involved in demand response to offer flexibility: Eventually, the INVADE platform aims to optimize the operation of batteries and loads to manage and allocate flexibility in order to increase the profit of the stakeholders. The stakeholders considered are the Distribution System Operators (DSO), the Balance Responsible Parties (BRP) and the Prosumers which all seek a profit from such flexibility service [13].

As a result, the optimization algorithm will schedule the operation of batteries, generators and loads with respect to each other taking into account the market and the power system. The constraints of reliability will be ensured in the model to create a useful and implementable result. The algorithm reacts to variations in the load and generation and responds quickly to imbalances before other type of control is needed. The result for the power system is an improvement on the quality and reliability of the energy generated and delivered, minimization of costs for the stakeholders and ultimately successful integration of RES [14], [13].

1.3. Research Setting

The eleven different partners of the project will research new ways to comply with the objectives mentioned in the previous section. The partners, shown in Figure 6 are different stakeholders: CITCEA, a research center from the Universitat Politècnica de Catalunya; Estabanell Energia a Spanish electric utility company; eSmart systems an IT-system company based in Norway; Lyse, a Norwegian industrial group; NTNU the Norwegian University of Science and Technology; Schneider Electric: a multinational company specialized in energy management and automation; ElaadNl, a Dutch innovation center for charging infrastructure; VTT: Technical Research Center of Finland; Greenflux: a company based in the Netherlands for charging infrastructure for EV; Albena: a Bulgarian brand in the sustainability Hotel line; and Smart Innovation Norway who hosts the Smart Energy Markets as a cluster of industry and academy.[15]

Figure 6. The eleven partners of the INVADE project

In this line, there are five pilot projects being developed [14]:

1) Norway: The pilot consists of taking advantage of the economic incentives offered to the consumers for flexibility and the area of opportunity is that Norway is the European country with more density of EV.

2) Germany: The project will integrate RES and batteries in two levels, community and household using the existing infrastructure and integrating ICT tools.

3) Spain: the pilot is aimed to demonstrate centralized storage as a safe and reliable alternative for the power system.

4) Bulgaria: the project focuses in assessing a centralized storage for supplying two hotels, restaurant, spa and swimming pools.

5) The Netherlands: this project consists in analyzing different approaches to charge EV using RES

This work was developed within the INVADE project in the facilities of CITCEA at the UPC in Barcelona. This research center is specialized in design and development of prototypes, especially focused on industry and power system applications. The main fields of research are: modern power systems, power electronics, RES, digital energy systems, Smart Grids and Energy Economics. [16]

Furthermore, this project was started in February 2018 and will be presented to an evaluation committee in October at UPC. It was conducted as the Final Master Thesis for the EIT InnoEnergy MSc. Energy for Smart Cities.

2. Introduction

The present work proposes a strategy that strongly aligns with the first specific goal of the INVADE project. Until now, a flexibility management system has been designed by modelling the different HEMS components and doing cost optimization within the specific electric markets of each pilot project.

The grid limitations are being taken into account as constraints in the model. The variability of the different components is tackled using the flexibility provided by the same or different components, especially storage. Finally, the uncertainty of the RES, EVs and other actors is being incorporated through predictions. However, the complexity of these predictions grows with the number of variables involved and its accuracy decreases.

2.1. Objectives of the Project

The aim of this project is to assess another alternative for coping with uncertainty: SP. Therefore, the main objective is to create a stochastic optimization for a Home Energy Management System within the INVADE project.

The specific objectives are:

- -Compare the stochastic and the deterministic approaches for the same model
- -Establish a method for quantifying the flexibility that can be obtained and its value
- -Use open software for all the stages of the optimization involved
- -Assess the environmental impact of the project

2.2. Relevance

Practical relevance

As previously mentioned, this work was developed within the INVADE project. It introduces a new possibility for the project to tackle its objective of dealing with uncertainty. The pilot projects give a possibility of direct application and benefits and it builds on the work of the different partners throughout Europe. The possible applications for SP are endless and it can be extended beyond HEMS to industrial energy management systems, EV charging spots by RES, etc. This thesis results open a new pathway for the INVADE project. Furthermore, it gives insight on the advantages and disadvantages of this method.

Scientific relevance

SP is still immature, however the benefits that can be obtained are clear and the barriers for implementing it are being lifted. Furthermore, it is a research undergone using open software which opens the possibilities for future researchers on the topic. The extension for SP used in this work – PySP- is a new tool which is still being developed. In the future, the evolution of this extension opens the field for more complex analysis.

2.3. Scope of the Project

The work presented here is an algorithm that optimizes the state of a prosumer as to minimize its costs during a day. As this thesis, is a first approach to SP, the algorithm optimizes a simple case of a prosumer which consists of only a PV panel generator, a battery and an inflexible load. However, the code is written with the possibility to add more components.

The approach used to assess uncertainty for HEMS is Stochastic Optimization. Consequently, the input data is just example data and is not appropriate for inferential statistics. The statistical analysis done just describes the behavior of the data in order to have an appropriate input for the model.

The present work only considers a two-stage stochastic problem. Even though it might be interesting to extend the research into more stages. The algorithm created supports multistage optimization.

2.4. Research Questions

The main research question for the project is: **"Which are the advantages of stochastic optimization over deterministic optimization for HEMS?**"

In order to answer this question, the problem is decomposed into the following sub questions that contribute to the objective of this work.

What are HEMS?

To have a proper context of the application it is necessary to understand what the current state of HEMS is as well as the different components which they are made up. Furthermore, it is necessary to see how the HEMS components are model within the INVADE project.

- Definition of a HEMS?
- Which components are part of a HEMS?
- What are the models for a HEMS and its components within the INVADE project?

What is Optimization?

To have a proper understanding of the purpose and value of this work, the first step is revising the optimization concept and understand how it is applied. Furthermore, the current state of optimization for HEMS is analyzed.

- Definition of optimization
- Which types of optimization are there?
- Optimization for HEMS

What are the most important characteristics of SP?

This work focuses on SP; therefore, a deep analysis should be done. The conceptualization, advantages, challenges, characterization of these problems.

- Definition of SP
- What are the necessary inputs and expected outputs of a SP algorithm?
- How are SP problems model and solved?

2.5. Thesis Structure

The present work has the following structure:

Chapter 1 gives a motivation of the field of research as well as what is the origin of the project and its research setting.

Chapter 2 sets the objectives, relevance and scope of the project, as well as the research questions to be answered.

Chapter 3 introduces the reader to Demand Response and gives a general overview of HEMS, its different components, as well as presents other approaches existing in literature to model it. Moreover, it aims to explain why HEMS are needed

Chapter 4 introduces optimization. Particularly, the presence of uncertainty in optimization is overviewed. SP as an alternative for coping with uncertainty is presented.

Chapter 5 details the methodology and steps followed to conduct the present research and the creation of the algorithm.

Chapter 6 presents the data obtained with the algorithm applied to a case study and the results to the specific objectives.

Finally, the thesis is concluded, drawing the main learnings and discussion points of this work.

3. Demand Response

Flexible refers to that which is "*characterized by a ready capability to adapt to new, different, or changing requirements*" [17]. The power system is instantly changing, and the system needs to react appropriately as fast as possible. The new characteristics of the power system require to find innovative strategies to increase flexibility. These have created new actors within the system such as HEMS that are part of a new concept of grid and it is at the center of this project.

3.1. Changes to the Grid

RES rely on the natural energy flows of each resource, therefore the technology must be located at or near the abundance of such resource, so the collection and production of energy is done on site. Furthermore, the output energy is normally variable and unpredictable due to the nature of the primary resource. Consequently, investment in infrastructure and important changes in the operation of the whole electric system have been happening and represent a technical and economic challenge. For example, many RES have been incorporated to the power system in a traditional, large, centralized manner. However, some technologies like PV panels can be installed at the point-of-use.[18]

These changes have contributed to the evolution of the conventional structure of a power system as it is shown in Figure 7. PV panels owned by residential users are now generating in the low voltage end of the system as opposed to the traditional generation in the high voltage end [8],[18]. This new type of user who consumes but can also produce energy is called a prosumer [19].

A smart grid, incorporates these changes and uses information to improve and optimize the operation of the system in order to become a fully automated and reliable network [20]. In general, the term "smart" refers to a device which can collect, send and/or react to information.

In recent years, new smart devices have been incorporated to the grid that enable bidirectional communication. Based on Home or Local Area Networks, it is possible to implement: advanced metering infrastructure, to monitor and send consumption data remotely; smart sensors, to collect and send the consumption data; smart home appliances, to control the appliance given a remote signal; etc. This is the foundation for HEMS [20].

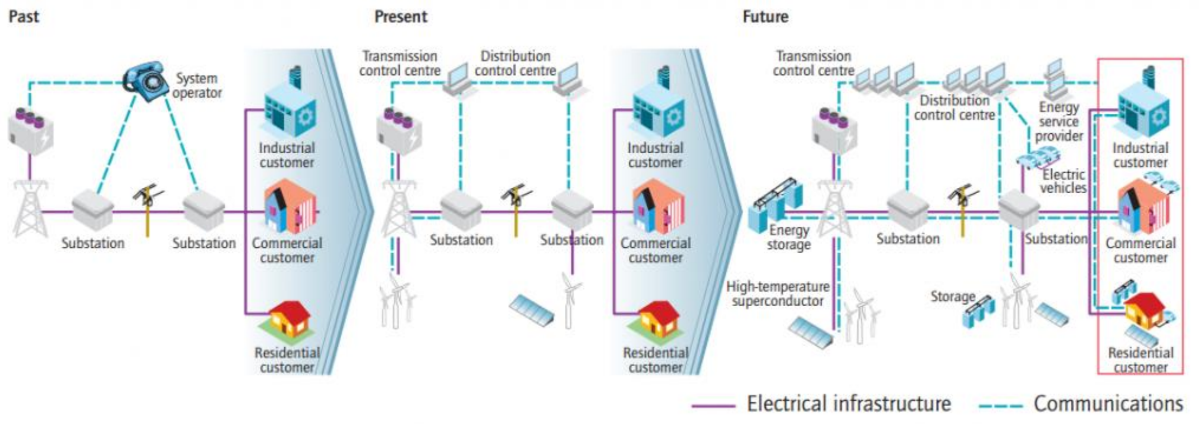

Figure 7. Evolution of a centralized electric grid into a decentralized smart grid [21].

HEMS are essential to use demand side management in the residential sector [20]. In [1] it is defined as "demand response tools that shift and curtail demand to improve the energy consumption and production profile of a dwelling on behalf of a consumer. HEMS usually create optimal consumption, production [and/or storage] schedules by considering multiple objectives such as energy costs, environmental concerns, load profiles, and consumer comfort."

From this definition, the components of a HEMS can be inferred. First, the infrastructure that allows bidirectional communication and control such as smart metering and sensors as well as the appropriate network. Second, at least some of the appliances and generators should have the necessary power electronics to be able to react to the HEMS. Third, the appropriate market signals that incentivize consumers to participate by offering flexibility. These price signals should be a reaction to the instant requirements of the grid. A HEMS then is able to shift and curtail the demand according to variable prices and comfort settings. A diagram for a HEMS is shown in Figure 8.

Within the project, it is assumed that the correct infrastructure is in place and the focus is to model the different components and appliances such as generators, batteries and loads. For this, the different characteristics of the components, as well as the possible degrees of control or "smartness", are considered and incorporated into the model.

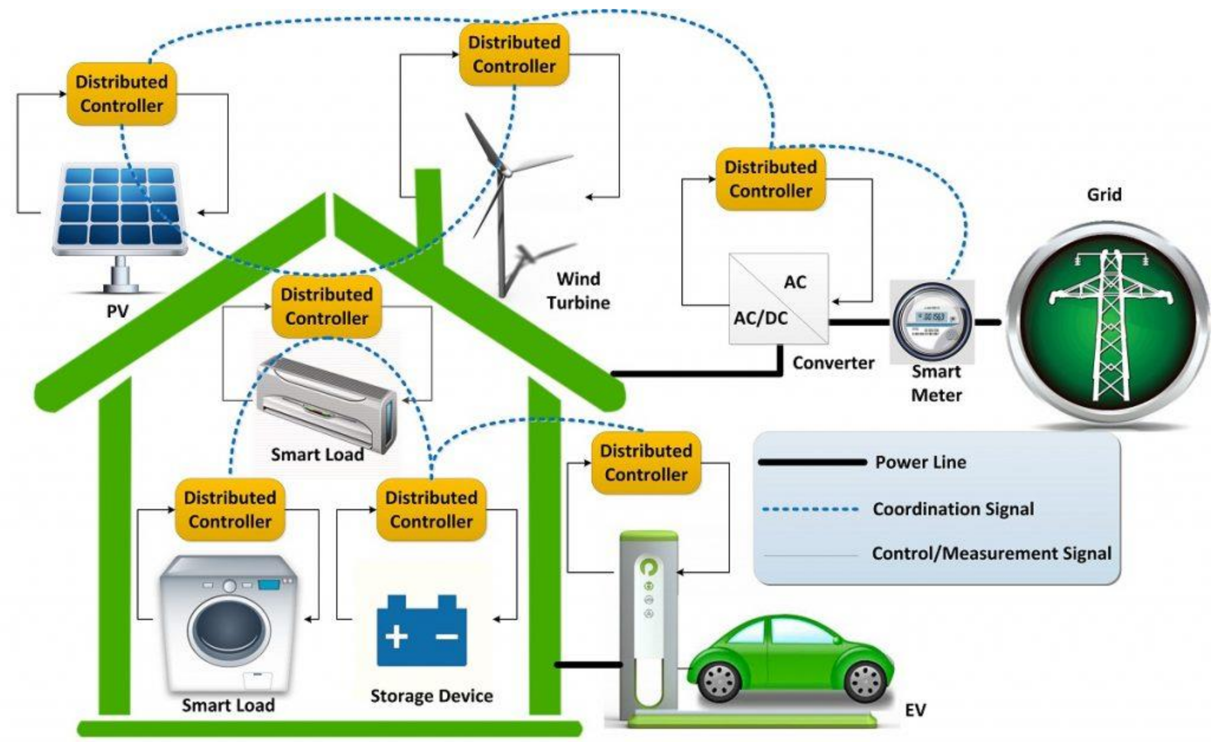

Figure 8. HEMS diagram [22]

3.2. Generators

A generator is characterized for injecting energy into the system. In a microgrid, these sources can be PV panels, small wind turbines, micro combined heat and power (CHP), or diesel generators. There are several types and classifications of these generators.

For the INVADE project there are:

Inflexible

This type of generation needs to be consumed when it is produced. Therefore, the scheduled production equals the predicted one.

Curtailable Reducible

Refers to those generators that can be modulated and controlled with power electronics to give a desired output are called reducible. These can range from zero to the predicted value.

Curtailable Disconnectable

Refers to those generators which either produce what is predicted or are disconnected.

From the different types of generators for microgrids, PV panels are the most common within HEMS. Depending on the type of power electronics of the system one can have a curtailable disconnectable or reducible generation unit [23].

More complex models like the one in [24] include the solar irradiation and the temperature of the cell to get more accurate predicted power. The output power is proportionate to the solar radiation and inversely proportional to the cell temperature. However, this is out of the scope of the INVADE project since the predicted PV values are taken as given.

Other type of generators like a CHP proposed by [25] and [26] are not yet incorporated into the INVADE project or to this work.

3.3. Batteries

Batteries are a great source of flexibility because they allow to shift energy generation and consumption over time. Therefore, the strict reliability constraint that supplied power and demanded power must be equal at all times can be fulfilled with less difficulty.

The State-of-Charge (SOC) refers to the amount of energy in a battery on a certain period t. This is the main variable to schedule in the HEMS and therefore has been modelled vastly.

A simple model, without losses, can be found in [24]. Other models -[19] and [27]- specify whether the energy comes from the PV or the grid. Furthermore, the variables can be defined with power, like in [19], [28] or with energy, like in the INVADE project.

Due to the sensible physical nature of batteries several constraints are needed to avoid damage.

State-of-Charge Range

One of the most common constraint ([19] [23] [27] [28] [29]) is about the total energy to charge or discharge the battery. Deep discharge or overcharge of the battery can greatly reduce its lifetime. Therefore, the SOC stays within certain limits. These limits are dependent on the nature of the battery and are given by the devices' data sheets.

Charging and Discharging Power Constraints.

As the previous constraint, this one can be found in most of the literature ([19] [23] [27] [28] [29]). [28], [19] limit the maximum charging and discharging power: if they are too big then the lifetime of the battery gets shorten. Therefore, they will range from the minimum to the maximum allowed. These constraints are normally given by the devices' data sheets.

The INVADE project limits the charging and discharging levels by the amount of energy consumed over a whole period t, this, depending on the pilot project, should consider not only the data sheet but also the peak powers contracted. Furthermore, it shapes the charging and discharging power curves according to the SOC to reduce degradation of the battery.

Charging and Discharging Simultaneity

[28], [30] propose to explicitly exclude the possibility of charging and discharging the battery at the same time to protect the device and ensure efficiency. The INVADE project does not have this implemented, as it is considered redundant and increases the number of binary variables in the model which increases complexity.

Battery Energy Balance

Since the batteries shift energy consumption in time, an energy balance is needed to ensure the conservation of energy [24]. This constraint ensures that the energy discharged cannot surpass the energy charged [27]. As well, an overall balance is done between load, battery, generation and grid [24].

In [29], is stated that the battery should not export electricity to the grid whereas [19] states that the battery can feed either the grid or the building. This consideration depends on the application.

Planning Horizon Constraints

There are more constraints such as equivalent-full-cycles or full-day-autonomy explored by [29], which aim to improve the performance of batteries in the long term. These constraints aim to have a certain amount of energy by the end of the planning horizon. This is especially relevant for EV batteries or off-grid batteries. For the INVADE project, a certain SOC is needed by the end of the planning horizon, but for this thesis it was not implemented.

Battery losses

These losses have been divided in two main groups, the losses caused by the battery float charge and the losses with regard to the efficiency [31].

The losses regarding the efficiency are taken into account by the INVADE project, both the charging and discharging efficiencies are considered while modelling the SOC. The battery float charge losses or self-discharge losses are neglected. The batteries modelled in the INVADE project are daily rechargeable batteries, therefore the self-discharge can be disregarded.

3.4. Miscellaneous Loads

Loads in a house are versatile and each has different characteristics. According to their nature, they can provide some or none flexibility. Appliances such as refrigerators, whose consumption should always be met, are called inflexible loads. Furthermore, thermal related equipment provides a wider range of flexibility through the mismatch of thermal and electrical transients.

3.4.1. Inflexible Loads

This type of load does not allow for any control, the energy demanded has to be met precisely at each period. These are called inflexible by the INVADE project but other names are given such as uncontrollable [23], essential [30] or must-run [32]. Fur purposes of the INVADE project and this work the inflexible loads are modelled altogether, however some authors like [30] model common loads like refrigerators separately.

3.4.2. Curtailable Loads

This type of loads allows its consumption to be reduced, either partially or totally. For curtailable loads, the model establishes the decision criteria to whether provide the energy forecasted or not in a specific period. For curtailable reducible loads, the energy delivered has a range of operation. For curtailable disconnectable loads, the energy delivered is either the nominal power or off.

3.4.3. Shiftable Loads

This type of load gives the opportunity to control when and sometimes how the energy is consumed. These are called shiftable loads by the INVADE project, but other names are given such as schedulable by [23]. There are different categories of shiftable loads.

Shiftable Profile

This type of units will only advance or delay the consumption profile. Therefore, the cost associated to it is proportional to the number of periods shifted.

Shiftable Volume

In this case, the load units are able to reduce, increase, advance and/or delay consumption. This provides great flexibility.

Shiftable Phases

A third approach done by [30] is phase shifting for those loads which operate in phases like a washing machine and it delays the phases when requested.

3.4.4. Thermal Loads

These differ from the electrical loads because one must take into account the thermal transients. Since the thermodynamic behavior has a different and slower transient than the electrical, there can be time mismatches of operation and consumption and gives flexibility opportunities. This is considering by modelling the thermal behavior of the building or house to maximize the benefits.

An example is an Electrical Water Heater (EWH). The main model follows the equivalent thermal circuit [26] follows a one-node model, which means the heat is distributed equally inside the container. Even though the EWH cannot be curtailable in its power, the user comfort has a wider range of operation, which gives the opportunity to have a controllable load [28] goes a step further and models the thermal resistance.

[23] presents a more advanced model and considers a thermal equivalent circuit of two nodes. In this case, the two-node model captures the stratification phenomenon inside the water tank which is a more realistic approach.

[23] and [30] developed an equivalent model for HVAC systems.

For this project, thermal loads are not incorporated.

3.5. Electrical Vehicles

Electrical Vehicles are equipped with a battery that can serve as a source of flexibility in the same manner as a regular battery. Depending on the type of controllers of the Electrical Vehicle, the amount of flexibility and its model differs greatly. However, since the EV is out of the scope of the project, this model is not further explored.

3.6. Energy Balances

An energy balance is an important constraint. Each balance depends on its components and it describes the power flows of the system. As it will be explained later, within the HEMS the power entering the system must equal the power exiting the system at all times.

$$P_{in} + P_{battery}^{discharge} = P_{out} + P_{battery}^{charging} \quad \forall\, t \in T \qquad \text{Equation 1}$$

However, as stated before in the batteries' constraints, batteries only shift energy within time, therefore, the discharged energy cannot exceed the charged energy.

$$E_{in} = E_{out} \qquad \text{Equation 2}$$

No Grid

Without a grid, the balancing power relies on the battery and the on-demand generators (whose output can be increased or decreased as will). The general off-grid power balance, as is given by [23] and [31], states that, when the energy produced by the PV exceeds the load, the battery should be charged, when the load exceeds the PV generation, the battery should be discharged.

Grid Connection

In a grid connected system, the balancing can be done by the grid, a battery or an on-demand generator. This is by far, the constraint which varies the most in the literature due to the variety of applications. [32] presents a balance of a system that has PV generation, no battery and the grid allows injection of energy. [23] and [27] present a balance with a similar system with a battery. [19] adds the balancing properties of an EV. [30] presents a detailed balance of the different loads specified.

Component Balances

It has been mentioned that batteries have normally a constraint that describes its own energy balance through time. Some models require that other components have balances as well. For example, in [23] there is a load balance that considers all the different types of loads in a HEMS. [19] and [30] do a balance on the PV.

Constraints

The most common constraint has to do with the physical and economic constraints of the components and transmission lines. In general, [24] states that both the power and the energy must be less or equal to the rated values of the components. Specifically, electricity bought or sold must be below the import and export capacity limits, respectively [13].

[13] remarks that, it is priority to respect the physical constraints for protection. In some cases, and due to the complex nature of the power system and the electricity market, the limit might be dynamic and time-dependent. In those cases, a time index is needed.

Simultaneity of importing and exporting energy from the grid is normally limited. This is mainly due to physical constraints, but is also relevant in the market. [19], [30], as well as the INVADE project, use integer variables to describe this constraint.

The energy balance constraint ensures that the optimization algorithm helps to maintain the reliability of the grid, which is the main issue to tackle and mitigate in order to have a successful migration to RES.

3.7. Reliability of the Grid

As stated, the introduction of RES is an important strategy to decarbonize the electricity system. However, the crucial challenge is to maintain its reliability. The INVADE project follows the opportunity that HEMS give, to maintain and even increase such reliability.

For a power system to be considered reliable these constraints must be met:

1) The power supplied, and the power demanded are exactly met at all times, in other words there should be a balance between what is being produced and what is being consumed at every single moment. Figure 3 makes a representation of this using a scale.

2) The generation capacity installed should be sufficient to meet the peaks of the demand. This is called capacity adequacy, and in other words, the maximum generation possible

should be more or equal than the maximum demand requested from the consumers.

3) The infrastructure for transporting and distributing energy must have the sufficient capacity to deliver the energy generated to the end consumers [7].

Less flexibility from the generation side, elevated costs due to RES technologies and distributed generation: The integration of renewables poses challenges to keep these constraints. The system needs higher resilience and flexibility to compensate for the intermittent nature of the RES as well as the location of such generation. Traditionally, demand is an inflexible load, in other words the amount of energy demanded cannot be controlled or influenced by the power system it is given by the users. The system balance is kept by adjusting the supply to meet the load, these sources are therefore called dispatchable. This is exemplified in Figure 9. In this case, an increase in load is met by an increase in generation, which for fossil fuels like coal meant increasing the amount of fuel put into the generator.

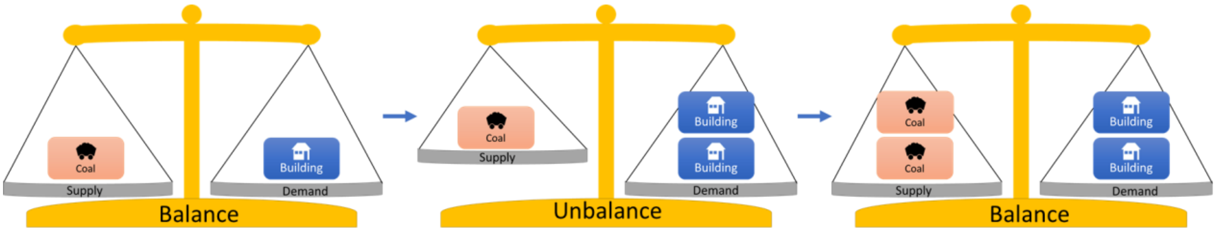

Figure 9. Representation of the traditional system balancing by adjusting the supply

In contrast, keeping the balance between a time-variable source and an inflexible load poses a challenge. As shown in Figure 10, balance can only be obtained when the supply matches the demand. On one hand, the supply can change and generate more (or less) energy than what is needed. On the other hand, it is also possible that the load will increase (or decrease) but there is no possibility to control the amount of sunlight on the supply side. In any case, the reliability balance constraint is not fulfilled. Therefore, it is expected to have an increase in the costs balance the system [18].

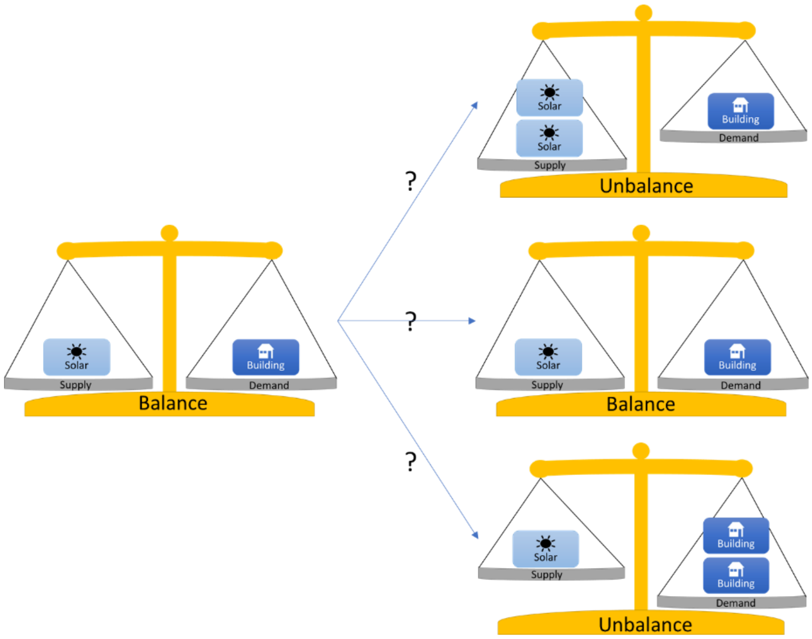

Figure 10. Representation of the power system imbalances due to variable generation and inflexible load

Another consequence of the volatile nature of RES involves generation capacity adequacy. As stated before, the maximum possible generation needs to be more than the maximum peak demand. However, with a variable RES the systems are oversized to compensate for a lower power output. For example, if the peak load of a house is of 1 kW at 8.00, a PV panel would need to be bigger than 1 kW in order to compensate for the fact that at 8.00 the Sun will not produce at 1 kW, but less [18].

Finally, RES are normally geographically distributed according to the available potentials. Therefore, investments are needed not only on the technologies but also in how to connect the generators to the system [8]. Furthermore, there needs to be adjustments to match the directions and magnitude of the power flows. As stated before, distributed generation in medium or low voltages is changing the dynamics and structure of the power system. The need of increasing the electric infrastructure is dependent on the location of the generation and the consumption and time mismatches. As well, different controls and equipment is needed to ensure the performance and the quality of the energy delivered. [8], [33].

In conclusion, the variability of RES increases the needs to keep the system balanced, creates a need of oversizing the capacity to meet demand peaks for correct capacity adequacy, and finally poses stress to the traditional power flow of the transmission and distribution system. Higher penetration of renewables will result in need of additional flexibility. Some strategies to bring such flexibility are:

1) Curtail generation in case there is extra generation due to low demand. However, in this case resources are not used optimally, but is easily implementable [34].

2) Use of storage for both saving the extra generation or withdraw energy from such storage. In this way mismatches in time can be addressed, reducing the impact of RES fluctuation. This allows for a more optimal use of resources and helps increase the reliability of the system. The drawback is that storage technologies are normally expensive and tend to rise the cost of energy from RES [18], therefore huge investments are being done in order to pursue a wide range of storage technologies. Storage is also useful to increase the capacity adequacy of RES [8].

3) Combination with other dispatchable existing plants and baseload technologies. This has the advantage that it diversifies an electricity mix and therefore increase its resilience. However, technologies with low ramping rates like nuclear or lack of heat storage for a CHP technology present further challenges [35]. If the flexibility comes from fossil fuels then the mitigation benefits of using RE are reduced [8].

4) Demand response aims to change the inflexible nature of the load, it is defined as "*load management triggered by power price signals derived from the spot market prices or other control signals*" according to the IEA. It has potentially low costs although the potential is somewhat limited [36]. Technical limitations exist as there is a need of a smart grid, with remote control and metering.

The INVADE project through the optimization of HEMS tackles three of these strategies, optimizing the system to reduce costs. In the end, the strategies adopted to mitigate of GHG emissions are a balance between environmental and economic considerations. The integration of RES into the system have several costs that need to be considered and are dependent upon the technology and the state of the system. Variable RES have higher balancing costs due to the required extra flexibility, higher capacity adequacy costs due to the oversizing of the system in order to cover peaks and higher transmission and distribution infrastructure costs due to the locations of the generation[1] [8].

---

[1] For example, Wind has a balancing cost that ranges from 1-7 USD/MWh, the capacity adequacy

The optimization algorithm presented in this work, as well as those being worked on within the INVADE project aim to ensure the cost-effectiveness of Demand Response for HEMS. The components of the system, the grid constraints and the requirements for reliability are integrated into a model that aims to schedule the operation of the components in the most cost-effective way.

Cost ranges from 0-10 USD/MWh and the electrical infrastructure cost ranges from 0-15 USD/MWh. PV balancing cost 1-7 USD/MWh, electric infrastructure costs from 0-15 USD/MWh but capacity adequacy costs are undetermined, due to the dependence of the location. Because of the coincidence of generation by PV and air conditioner usage, PV can produce savings instead of costs for capacity adequacy of around 23 USD/MWh [18], [33] (The data is from 2011, the exchange rate was of 1 USD = 0.772 Eur)

4. Optimization

According to the Merriam Webster dictionary Optimization is “*the act, process or methodology of making something as fully perfect, functional, or effective as possible. Specifically, in mathematics, optimization refers to the mathematical procedures [to achieve what is stated before]- such as finding the maximum of a function”* [37]. The INVADE project is developing the methodology to operate the components of a HEMS in the most functional way as possible, by considering cost minimization.

Optimization is one of the oldest branches of study in Mathematics. Its creation helped develop geometry and differential calculus. Since then applications for optimization have become more popular, especially in the fields of science, engineering and technology [38].

When a problem is encountered, it is common to try to find solutions for it. However, many times it is not about finding a solution but finding the best solution possible according to certain criteria. The process of finding such solution is called optimization [39].

Often this process is done implicitly by using common sense, references, opinions or one’s own experience. However, this complicates further when dealing with many decision variables or constraints. Furthermore, intuition does not offer the certainty that the solution found was the best one [39].

As previously mentioned, optimization is used to solve a range of problems. Consequently, there are different methods and different types of optimization which should be carefully chosen to achieve the desired results [38]. For HEMS, the different modelling approaches as well as the particular components of the system, the parameters and the objectives of each application result in a variety of methods for optimization. In addition. the main challenges to consider are the forecasts’ uncertainty, the devices heterogeneity, multi-objectivity, computational and timing limitations and the inclusion of the consumer’s well-being [1].

For this thesis, the algorithm schedules the operation of the batteries with respect to the expected generation and consumption. The aim is to minimize the cost for the house. It is a multivariable, multi-objective, mixed integer linear program (MILP). These characteristics are described next.

A multivariable optimization deals with more than one design variable [39]. In the INVADE project there are several design functions per component. In this project’s scope, there are thirteen design variables which will be later specified.

A MILP is an optimization that requires that some of the design variables are integers and that the objective and constraints are linear functions. Also, the constraints need to be non-negative [40]. A linear program is easier and faster to solve, especially when considering the discrete nature of some variables. Such variables increase the complexity of the optimization. Whereas in some cases design variables could round up or down, in others this would result in inviable and non-optimal results [41]. The INVADE model uses binary variables to represent the on-off state of the devices, as well as the sell-buy state of the system. Within this project only two binary variables are used. As well, the functions and constraints have been modelled to have a linear program.

Even though one-objective optimization is the most common, sometimes is of interest to have more than one objective into consideration. However, the sometimes conflicting nature of the different objectives make it impossible to have more than two or three [1], [39]. In the INVADE project, there are different interests according to different stakeholders using HEMS flexibility and therefore there are multiple objectives. For the prosumer, the main one is cost minimization -which is explicitly stated as the objective function. However, keeping the users' well-being is of utter importance as to ease the acceptance of HEMS in the residential area. Both objectives are directly affected by the uncertainty of the decisions to take, which poses a risk for the system. This is being assessed with different strategies.

4.1. Coping with Uncertainty in HEMS

Risk is defined as the possibility of loss [42]. By taking decisions without an absolute certainty of the future a risk is associated to such decision, as a result of the possibility of losing value with a different outcome. Some of the information used to schedule the HEMS might not be known with absolute certainty, nonetheless, decisions need to be made.

Uncertainty is present in the weather, PV generation, market prices, load demand, the user's schedule, etc. There are different methods to do the forecasting of the uncertainty: neural networks, fuzzy logic, support vector machines. The incorporation of forecasting is needed to do an a *priori* optimization for the planning horizon and the results are strongly dependent on the amount of certainty of such forecasts [1]. These type of decisions -in which the outcome is uncertain- are made by evaluating the possible outcomes, mainly their desirability their likelihood

[43]. The expected value is then calculated as the different possible outputs weighted by the probability of their occurrence [43].

There are risk-averse decisions and risk-seeking decisions. In the first ones, there is a preference for a sure outcome even though there is a possibility of a greater value, whereas the second ones are decisions where it is preferred to gain value even though it is less certain. In general, it is said that humans are risk-averse beings [43]. In a HEMS, the risk of unbalancing the system, or failing to fulfill the user's demands have a cost that might be too high to ignore. Therefore, HEMS decisions are normally risk averse [1].

One approach to uncertainty is to try to directly reduce it or ignore it. For example, some HEMS assume that there will be a perfect forecasting [32] others consider a certain error within the forecast [44]–[46], in this case a deterministic optimization is used and considerable effort is done to increase its accuracy and minimize the error. Others like [47], [48], make real-time decisions. For human patterns such as schedules, machine learning can be used like in [49], [50]; however, this is not useful in the case of RES.

Another approach is to incorporate uncertainty into the optimization, which according to [1] has the potential to improve scheduling efficiency. Some techniques are model predictive control [51], [52], statistical decision theory, decision analysis, stochastic optimization [52], [53], Markov decision processes, robust optimization [53]–[55], stochastic fuzzy optimization, stochastic dynamic programming [56], among others. A complete review can be found in [1] and [3].

This thesis focuses on SP. Hence, the uncertain data is represented as random variables. These random variables can be described accurately by means of a probability distribution, density or measures [3]. For this thesis, generation and load are assessed, it is assumed they follow a normal distribution per hour. A deterministic and robust optimization are also performed to benchmark the SP algorithm developed.

4.1.1. Deterministic Optimization

The deterministic approach considers the future fully and perfectly known [3]. This is the current approach of the INVADE project. As previously mentioned, the unreliability is internalized in a prediction. Sufficient effort and resources are put into creating such prediction. Afterwards, the system considers that the prediction will be true and therefore all the parameters are known. Nonetheless, there are limits regarding the prediction's complexity and accuracy. Moreover, the optimal solution might be optimal only if the prediction is fulfilled. A multistage approach can help reduce such uncertainty, in the time of the decision-to-make is taken into account.

[57] uses the next equation to predict the demand for the day-ahead $\widehat{D}_t$. This equation considers the demand of the previous 24 hours $D_{(t-24)}$, the median demand of the previous week $\overline{D}_t$ and the predicted temperature $\hat{T}_t$.

$$\widehat{D}_t = f_1\left(D_{(t-24)}\right) + f_2(\overline{D}_t) + f_3\ (\hat{T}_t) \qquad \text{Equation 3}$$

The current model of the INVADE covers different types of generators, loads, EVs and batteries. The design variables are specific to the nature of each component, the objective is specific to each pilot and the constraints are component and pilot specific[2]. In this thesis, the basic Reference Model is the deterministic model. Furthermore, deterministic optimization is done by using the expected values or average values for the load and the generation. This serves as a benchmark for the stochastic algorithm as stated before.

4.1.2. Robust Optimization

The robust optimization will consider the worst-case scenario and minimize its impact [1]. In this sense, it reduces risk to a minimum. However, this strategy might be not optimal when the cost of loss of profit is greater than the cost of risk [58]. [58] uses a hybrid stochastic and robust optimization to deal with the uncertainty of the day-ahead market. The robust approach trades cost for risk minimization. In this work, a robust optimization is done to compare the added value of stochastic optimization.

In this thesis, the robust optimization is done considering the worst-case scenario of the load- the consumption was the upper decile; and the worst-case scenario of the PV panel- the generation was as in the lower decile. In both cases, the mean of the data and a standard deviation of 1 was used.

4.1.3. Stochastic Optimization

In stochastic optimization the uncertainty of the unknown parameters is specifically taken into account while optimizing the design variables [3], [40]. The premise for using stochastic

[2] For the full model of components refer to [13]

optimization is that when dealing with uncertainty, it is impossible to find one ideal solution that meets all circumstances [3]. In general, this method offers a bigger flexibility to meet different outcomes. Scenarios of the possible outcomes are created, and each is weighted by its probability of occurrence. [40].

In [57], the uncertainty of the PV generation and the load in a HEMS was tackled using stochastic optimization and it reduced costs a 5.8% compared to the deterministic approach. As it can be seen in Figure 11, the energy purchased from the grid in the stochastic case is less than or equal to the energy purchased in the deterministic case. The stochastic parameters are calculated using quantile probabilistic forecasting

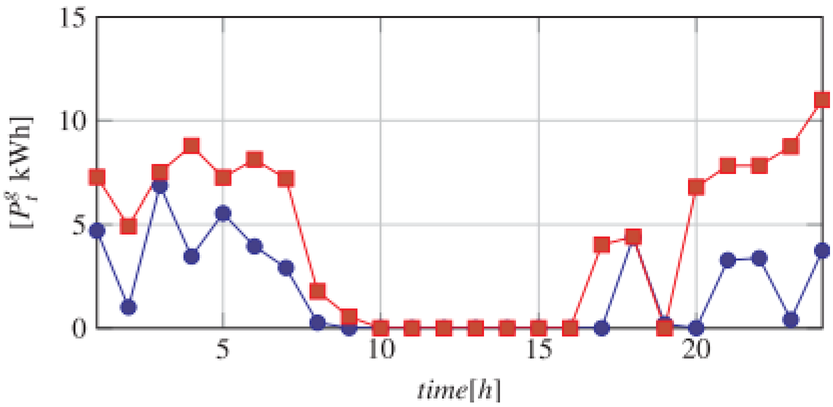

Figure 11. Comparison of deterministic vs. stochastic HEMS algorithms [57]

Furthermore, [57] concludes that the stochastic approach incorporates the robust approach and the optimal approach, increasing benefits in the long run.

A deterministic optimization is a particular case of a stochastic optimization where the probability of occurrence is equal to 1. Therefore, stochastic optimization is a generalization of problems where there is a lack of certainty regarding uncontrollable data [3]. Furthermore, a stochastic problem normally presents the following:

- Several decision variables with several potential values
- Discrete time periods where decisions are made
- The objectives include expectation functions or values
- The probability distributions of the unknown parameters are known

This work is a first approach of the INVADE project to stochastic optimization. There are several parameters that are uncertain, and their modelling becomes a complex task. For purpose of this first analysis two parameters are considered unknown: the generation and the inflexible load. Both have expected values and vary greatly during the day. Decisions with respect to the battery are done before the information is revealed. The objective cost function includes an expected cost for the future. Finally, for the scope of this project the probability distribution of the parameters is known.

4.2. Stochastic Programming

When the complexity of the problems to optimize increases it is normal to approach them using computers. In this way, an algorithm searches for the best choice of design variables according to the desired criteria as shown in Figure 12. The advantages are manifold: efficiency, time-saving, complexity.

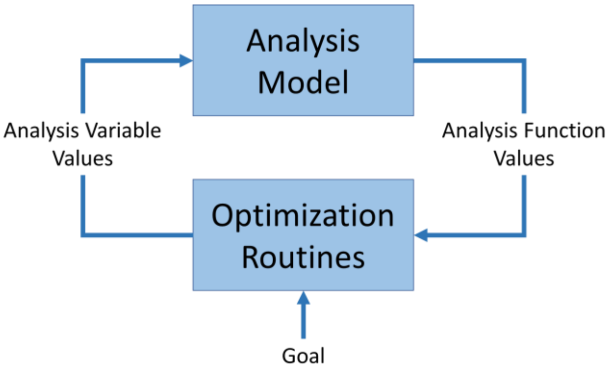

Figure 12. Block diagram of an optimization algorithm

There is abundant literature dating as back as 1955, for Stochastic Optimization, however, there was no possibility of solving the problems due to the complexity of the numerical methods needed to solve them [59]. The extensive form, later described, could be solved directly by regular optimization algorithms; however, this form represented a limit to the power of SP [3], [60]. Recently, a wide variety of algorithms are recently being applied to optimization problems to cope with uncertainty. Some of these methods are: Pure Random Search, Accelerated Random Search, Hybrid and Multistart Algorithms, Simulated Annealing, Genetic Algorithms, Markov Chain Monte Carlo, Particle Swarm Optimization, Ant Colony Optimization, Grenade Explosion Method, Differential Evolution, etc. [3], [59]. The one that was explored in this project was the

Progressive Hedging Algorithm by Rockafellar and Wets design in 1991.

The main general formulation for a two-stage SP is [3], [40]:

$$\max c^T x + \sum_{s \epsilon S} p_s d_s^T y_s \qquad \text{Equation 4}$$
$$s.t. \quad Ax = b$$
$$T_s x + W_s y_s = h_s, \quad s \epsilon S$$
$$x \geq 0$$
$$y_s \geq 0, \quad s \epsilon S$$

Where S represent the set of scenarios s, p_s the probability associated to each scenario. Decisions are separated into stages where x are first-stage decisions and y_s are second-stage decisions dependent on the scenario. Concepts such as extensive form, progressive hedging, stages and scenarios will be developed in the next section.

4.2.1. Definitions

There are several concepts regarding SP that need to be understood that are relevant for this project.

Stages

A stochastic problem is decomposed in stages according to the availability of information. Every stage corresponds to a period where decisions regarding an uncertain future are made, except for the last stage where all information is disclosed. For example, a two-stage problem will take decisions with respect to unknown information only once, and the second stage corresponds to the stage where all information is known. For a three-stage problem, the second stage will only reveal part of the information and new decision should be made with respect to the future, still the future is uncertain [3]. It is important to mention that the complexity of the problem increases exponentially with the number of stages [61].

As stated before, for this project a two-stage model was used. In the first stage, decisions over the battery are done regarding the uncertain future of the load and the generation. The first stage is prior to the beginning of the planning horizon, it is in $t < 0$, therefore the duration of the first stage is irrelevant. The second stage lasts the whole planning horizon, which for this analysis was of 24 hours. It starts since $t = 0$ as decision have been made and the information is being discovered. The batteries setup will not be changed as it was set in the first stage. In Figure 13, a two-stage SP is shown, whereas Figure 14 shows a three stage SP. Is worth noticing, that the difference of stages between these two diagrams, is due to the timing and type of decisions. In both, decisions are being made on a stage for which there is still not perfect information.

Recourse

It is a characteristic of some Stochastic Problems where corrective actions can be taken once the uncertainty is disclosed. The corrective actions are done with full information in each new stage [3], [62]. In this sense, a recourse decision is with respect to the present situation.

For this study, recourse actions are taken throughout the planning horizon. Every hour, information regarding production and consumption is being discovered and decisions with respect to the grid are taken. However, all the information needed to make these decisions is known. Both, Figure 13 and Figure 14, show a SP with recourse. In Figure 13, the recourse is done for the periods T=1 and T=2 and are dependent on the scenario. In the moment the information of the period 1 is known, recourse actions will be taken according to the scenario fulfilled. In Figure 14, the recourse is also done for periods T=1 and T=2, however in this SP these recourse decisions are dependent on the corresponding node.

Scenario

Scenarios help discretize uncertainties using a scenario tree. Each of them represents a possible situation or outcome after certain event has disclosed information. They have a probability associated to it [40]. There are several techniques to create the scenarios, however this is out of the scope of this work. Scenarios can be made up of a bundle of outcomes of a certain variable, as long as the bundle have the same impact or expected value [3], [62].

Within this analysis, 10 scenarios were generated per simulation. The scenarios had all a 10% probability and were obtained from the analysis of example data that follows a normal distribution.

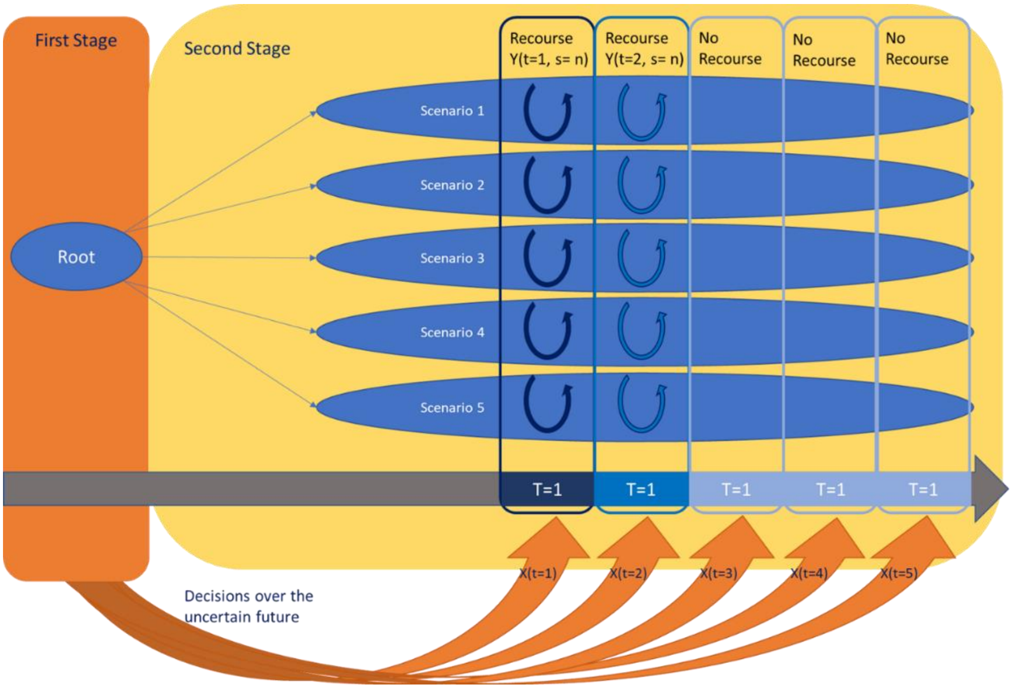

Figure 13. Representation of stochastic problem with two variables x(t) and y(t, s). Recourse actions over variable y(t, s) are taken in periods T=1 and T=2 according to the scenario uncovered. Uncertain decisions are all taken in the same moment, therefore it is only a two-stage formulation.

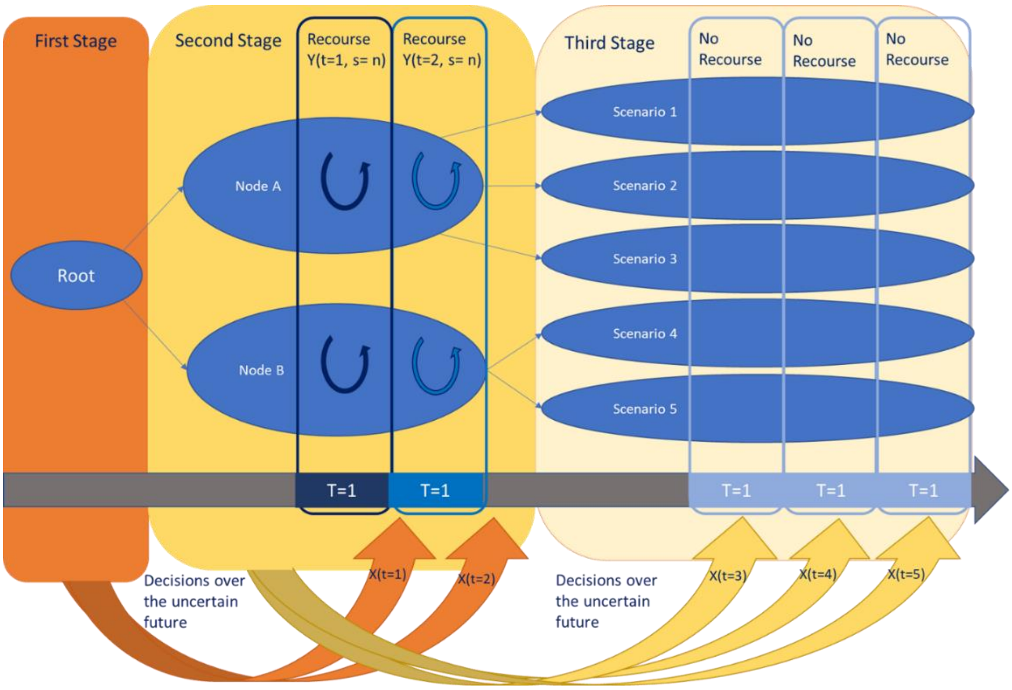

Figure 14. Representation of stochastic problem with two variables x(t) and y(t, s). Recourse actions over variable y(t, s) are taken in periods T=1 and T=2 according to the node uncovered. Uncertain decisions are taken in two different moments; therefore, it is a three-stage formulation.

Extensive Form

This is a possible formulation for the model of the problem. It explicitly describes each stage-decision variable for all scenarios. Furthermore, a decision second-stage vector is associated to the random vector according to its probabilities. It is the basic approach to stochastic problems and is easily solvable by classical algorithms such as L-shaped method [3]. This formulation has limitations in complexity and therefore it cannot represent all problems. It is also known as the deterministic equivalent and therefore, the standard deterministic solvers suffice. The stages are managed through coupled non-anticipativity constraints.

The extensive form is supported by the solver used, however it was only possible to use in simple cases with only one stochastic variable, therefore it was opted to use progressive hedging instead.

Progressive Hedging (PH)

This is a possible formulation for the stochastic model of the problem proposed by Rockafellar and Wets in 1991. It is classified as a horizontal decomposition strategy. The aim is to separate each scenario individually for each iteration. The method splits the problem into subproblems, finds solutions for each and then iterates until the solutions converge to the same value for the common nodes between scenarios [3], [63]. PH theoretically converges if all the decision variables are continuous [2].

The algorithm works as follows [3], [60],[2]:

1. For each of the 10 scenarios, solutions are obtained for the problem of minimizing the cost, subject to the different constraints. Each uses the deterministic formulation.

2. The variable values for an implementable-however not necessarily admissible- solution are obtained by averaging over all scenarios at a scenario tree node.

3. For each of the 10 scenarios, solutions are obtained for the problem of minimizing the cost, subject to the different constraints. Each uses the deterministic formulation, but terms that penalize the lack of implementability are added, using a sub-gradient estimator for the non-anticipativity constraints and a squared penalty term.

4. If the solutions have not converged sufficiently and the allocated computational time is not exceeded, the algorithm returns to Step 2 and iterates until one of these conditions is met. The solution found is implementable and admissible.

A block diagram representing the iterative process done by the PH algorithm is shown in Figure 15.

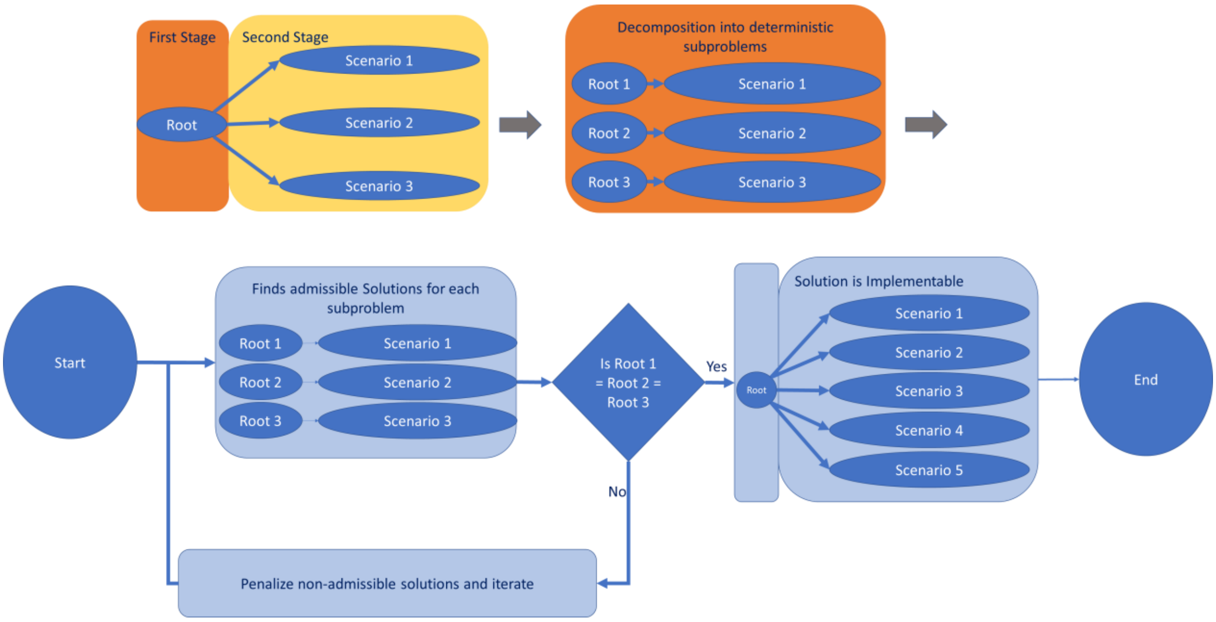

Figure 15. PHA representation, showing implementable and admissible solutions

Admissibility

It is the characteristic of a valid solution that fulfills all the constraints for all of the problems and subproblems [64], [63].

Implementable

In the case of progressive hedging, implementable means that all of the admissible solutions for each of the subproblems converge in those nodes that are shared by more than one scenario [63].

For example, in a two-stage problem with three scenarios, three separated subproblems are created and an admissible set of values is found for each subproblem. These solutions comprise first-stage variables and second-stage variables. If the first-stage variables of the three solutions converge, then the solutions are implementable.

Value-at-risk (VaR)/Conditional Value-at-risk (CVaR)

The VaR is a constraint established to minimize risk. It refers to the biggest loss of value-or biggest increase in cost- that can occur with a probability α. Mathematically it is defined as:

$$VaR_\alpha\big(q(\omega)^T y(\omega)\big) = \min\{t | P(q(\omega)^T y(\omega) \leq t) \geq \alpha\} \qquad \text{Equation 5}$$

Where $q(\omega)^T$is second stage data, $y(\omega)$ is second stage decisions, t is the loss of profit, and $P(q(\omega)^T y(\omega) \leq t)$ is the probability of the expectation of the second stage objective to be less or equal than t. When one wants to limit the losses, one can set the value of $\bar{t}$ with probability α [3].

Related to VaR, the CVaR also helps to minimize risk. Moreover, it satisfies a series of conditions to be a coherent risk measure whereas VaR might not always satisfy them[3]. It is defined as:

$$CVaR_\alpha(\xi) = \min_t t + \frac{1}{1-\alpha} E_{P_\alpha}[(\xi - t)^+] \qquad \text{Equation 6}$$

Where α is the confidence level, ξ is a random loss with distribution function P, and P_α is the distribution function:

$$P_\alpha(t) = \begin{cases} 0 & if\ t < VaR_\alpha(\xi); \\ \frac{P(t)-\alpha}{1-\alpha} & if\ t \geq VaR_\alpha(\xi). \end{cases} \qquad \text{Equation 7}$$

Expected Value

It is also known as Expected Monetary Value (EMV) [3]. The definition is a "*weighted average of the payoffs for a decision alternative* " [65]. For every possible scenario s, the output value R_{ds} is weighted with its probability of occurrence p_s. The sum of this weighted values gives the expected value per scenario. There is one value per decision alternative, in case of more than one decision alternative d, the maximum of these values-the best of these values- is the Expected Value EMV.

$$EMV = \max_d \sum_s p_s R_{ds} \qquad \text{Equation 8}$$

[3] For more information on coherent risk measures see Annex 1

Expected Value of Perfect Information (EVPI)

"*The loss of profit due to the presence of uncertainty*". It is a measure of the importance or the cost of knowing with certainty what will the future bring [3]. It is calculated by:

$$EVPI = WS - EMV \qquad \text{Equation 9}$$

Where EMV refers to the Expected Monetary Value, and WS refers to a wait-and-see value. The WS is an estimation of what would be the benefit if the decisions could be taken in retrospective, hence with perfect information about the future [3]. It is the best-case-scenario. For every decision alternative d, the maximum output value R_{ds} per scenario is weighted by the probability of such scenario. The sum of the different values for the different scenarios gives the WS.

$$WS = \sum_s p_s (\max_d R_{ds}) \qquad \text{Equation 10}$$

A small EVPI results in little profit from knowing the future with certainty.

Value of Stochastic Solution (VSS)

"The possible gain from solving the stochastic model." This value consists of the difference between the stochastic decision taken today and the expected value solution [3]. Whereas the WS considers a set of solutions, the EEV considers only one set of solutions- the deterministic model solution. EEV is calculated using the mean values instead of random.

$$VSS = EEV - EMV \qquad \text{Equation 11}$$

A small value of VSS denotes that the stochastic program can be approximated with mean values instead of random [66].

For an example on calculating EMV, EVPI and VSS go to Annex 2.

4.2.2. Requirements

In order to employ optimization, there are some requirements:

- Quantitative Model: it allows to compute or calculate the possible solutions. It needs to be a valid and accurate model of the problem.

- Design Variables: also called degrees of freedom. These are the independent variables which are used to make decisions.

- Design Functions: comprised by objective and constraints. The objective is the purpose or goal of the optimization, it is a minimization or a maximization function. The constraints are the limits within which the solution has to stay or the targets that need to be fulfilled, they are inequalities or equations.

Furthermore, for SP other requirements are needed to handle the uncertainties:

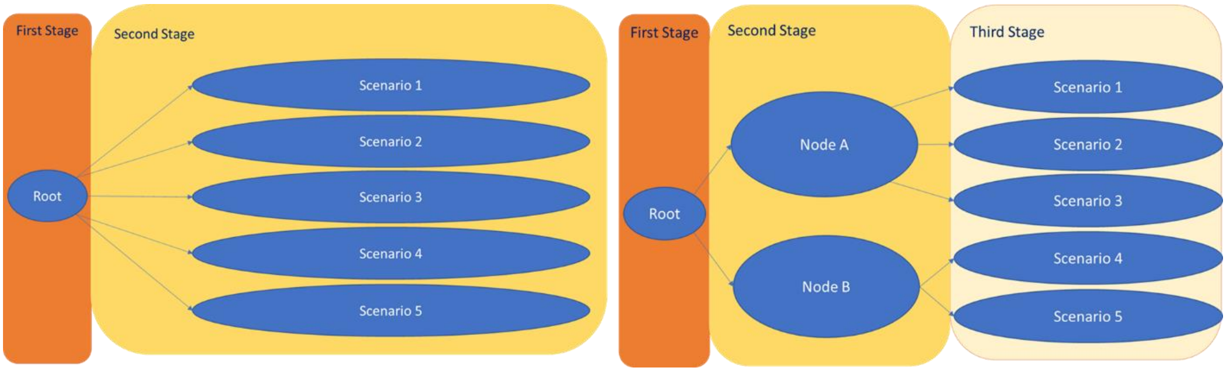

Figure 16. Two scenario trees. The first one is a two-stage, five- scenario tree. There are six nodes: the present reality and five possible future realities. Each arrow is associated to the probability of occurrence of such reality or scenario.

The second tree shows a three-stage, five-scenario tree. It has eight nodes, the present reality, two possible intermediate realities, and five possible final realities. Each arrow is associated to the probability of occurrence of such reality or scenario

- A scenario tree: it is designed to establish the relationship between the uncertainties, the outcomes and the time stages. The first step is to establish the number of stages and the number of nodes per stage. A node is a representation of reality. All have only one father but can have one or more children. A father node and a children node are never in the same stage. The first stage has only one node and is the only one without a father. Afterwards, the children nodes are connected to their fathers by probabilities.

Finally, processes, parameters and outcomes are assigned to each node and scenarios are associated to the nodes of the final stage [60], [62]. Two representations of scenario trees can be found in Figure 16.

- Data instance per scenario: For each scenario a possible realization of the future needs to be specified. Therefore, it is necessary to assign the corresponding data to the appropriate scenario [60]. In Figure 17, a diagram of the creation of instances per scenario is shown. A scenario instance encompasses the whole path from the root node until the final realization of such scenario.

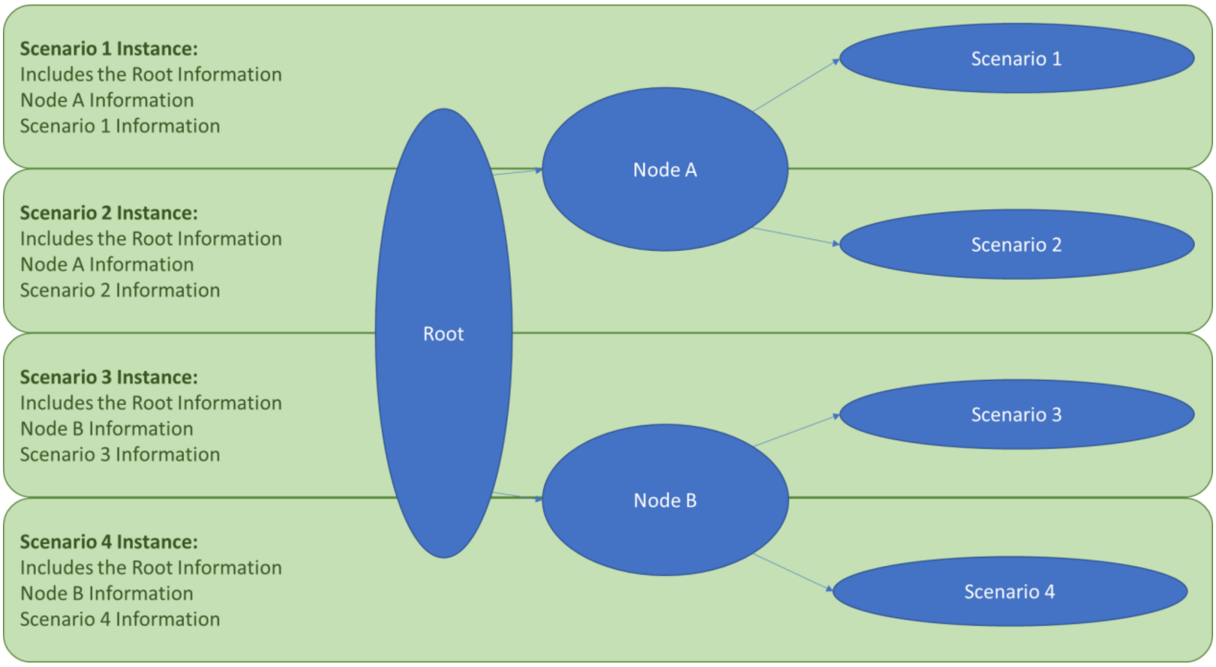

Figure 17. Data input per scenario instance.

- Data instance per node: Alternatively, it is possible to assign the corresponding parameters per node. This is preferred when the variations per scenario are small, and there are many common nodes as to avoid repetition of data. In Figure 18, a diagram of the creation of instances per node is shown. A node instance includes only the information related to that specific node, independent of the rest.

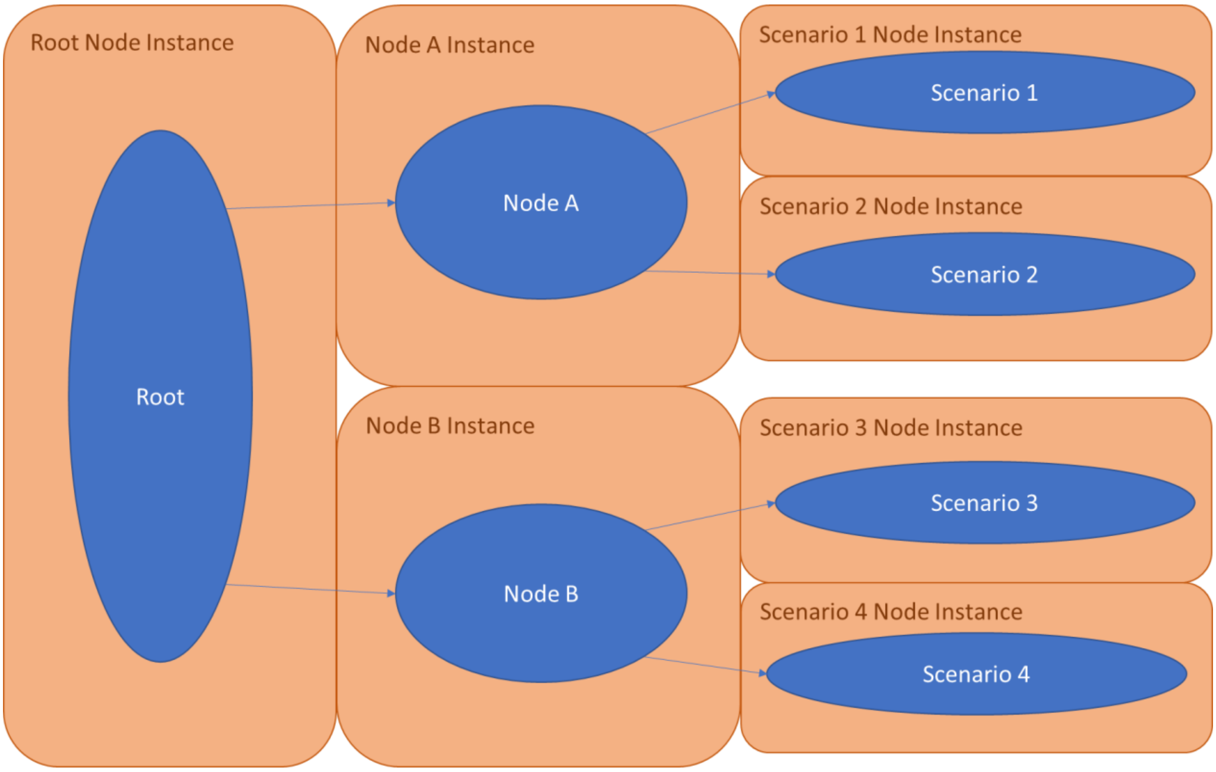

Figure 18. Data input per node instance.

After translating these requirements into a computer language, the software will call the model repeatedly while it finds possible solutions and evaluates them to find the optimal one. A computer has the advantage that gives a whole design space where it is possible to see different combinations and how changing the design variables affect the final solution [39].

The solution given by the algorithm is divided according to the number of stages of the stochastic program. For example, for a two-stage program, the first part of the solution will be the decision variables of the first stage and the cost of these stage. Then, the decision variables of the second stage as well as the recourse decisions and expected cost will be given according to each scenario. Finally, a probability weighted cost of the expected costs is given as a summary of the optimization.

5. Methodology

For this thesis, the quantitative model, design variables and design functions were created based on the INVADE documentation of the deliverable 5.3 [13]. Later, the implementation was done in Python using the Pyomo library. As first step, the deterministic optimization algorithm was done. Afterwards, the programming language R was used to do the statistical description of the input data. To create the SP algorithm, the Pyomo extension PySP was used and the scenario tree, together with the data instances gave the basic model of SP. A further step was taken with the libraries NetworkX and Pandas, to automatize the creation of the SP algorithm. In this way, it is possible to continuously update the input data.

An overview of the algorithm created can be seen in Figure 19.

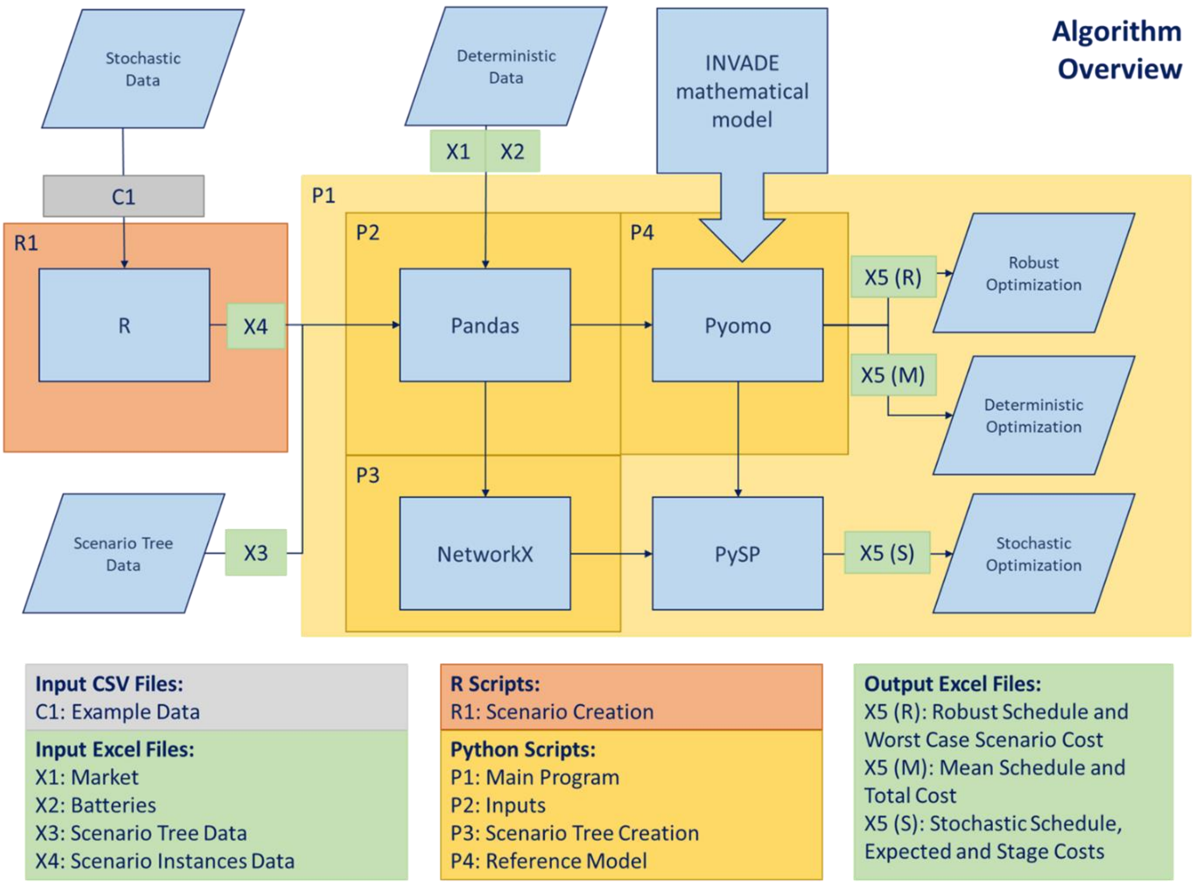

Figure 19. Block diagram of the algorithm created for this thesis

The algorithm consists of five excel files (X1-X5), one csv file (C1), one R script (R1) and four python scripts (P1- P4): Subsequently, the files and scripts will be referenced to the nomenclature given in Figure 19. For the parameter data, the stochastic data is read from C1 by R1, which in turn converts this information into the instance data X4. The deterministic data, consisting of the market information X1 and the battery parameters X2 is read from individual excel files directly with the Pandas library with P2. Finally, the scenario tree information is given as an excel file X3 and is processed with P2 to be used by NetworkX in P3, which in turn converts it into a Scenario Tree object that can be used by the PySP extension. P4 consists of the creation of the model object with the data provided by P2. The main program, P1, calls P2, P3 and P4, and uses PySP commands to solve the program and creates an excel file X5 as output. This output can vary depending on which case is being optimized. The robust optimization uses worst-case-scenario data creates X5 (R); the deterministic optimization which uses mean data creates X5 (M); or the stochastic optimization which uses different scenarios data and creates X5 (S).

The scripts can be found in Annex 3.

5.1. Mathematical Model

The model used for the components and for the HEMS are physical models. As mentioned, for purposes of this work only a simple version of a HEMS is studied. It consists of a single household with a curtailable reducible PV generator, one fully controllable battery, an inflexible load and the connection to the grid. For this work, a planning horizon of 24 hours was studied. It is assumed that the generator, battery, load and grid are connected to a meter and information on the energy consumption and/or production of each component is known. Even though only a household was studied, the algorithm was done with the possibility of scalability with more households and components.

In general, the load can be perfectly supplied with the grid and PV production. However, it is desirable to minimize the costs of the prosumer by adjusting the operation of the devices in reaction to the market tariffs.

5.1.1. Analysis Variables

5.1.1.1. Sets

Symbol	Explanation	Component	Value

T	Set of periods within the planning horizon. The periods are normally taken as hourly.	General	0-23 hours
T_{restr}	Subset of Set *T*. It includes all the periods on the planning horizon except t=0.	General	1-23 hours
B	Number of batteries in the analysis. Only one was considered.	Battery	1 battery

5.1.1.2. Parameters

Symbol	Explanation	Component	Value
P_t^{r-b}	Price at period *t* at which electricity is bought with a retail contract [€/kWh]	Grid	(0.06-0.13)
P_t^{g-b}	Price at which electricity is bought with a grid contract [€/kWh]	Grid	0
P^{tax}	Sum of all taxes that are related to the purchase of electricity [€/kWh]	Grid	0.1
P^{VAT}	Value Added Tax that is calculated over the total amount bought [%]	Grid	25
P_t^{r-s}	Price at which electricity is sold with a retail contract [€/kWh]	Grid	0.05
P_t^{g-s}	Price at which electricity is sold with a grid contract [€/kWh]	Grid	0
X^{imp}	Maximum import capacity from the grid [kW]	Grid	23
X^{exp}	Maximum export capacity to the grid [kW]	Grid	5
O_b^{min}	Minimum state of charge allowed for battery *b* [kWh]	Battery	0.5

O_b^{max}	Maximum state of charge allowed for battery *b* [kWh]	Battery	4
Q_b^{ch}	Battery *b* maximum charging power allowed [kW]	Battery	1
Q_b^{dis}	Battery *b* maximum discharging power allowed [kW]	Battery	1
A_b^{ch}	Battery *b* charging efficiency [%]	Battery	95
A_b^{dis}	Battery *b* discharging efficiency [%]	Battery	95
S_b^{ch}	Battery *b* charging threshold [%]	Battery	80
S_b^{dis}	Battery *b* discharging threshold [%.]	Battery	10
$\sigma_{b,0}^{soc-in}$	Battery *b* SOC at *t=0*, the beginning of the planning horizon [kWh]	Battery	1
$P_{b,t}^{ch}$	Battery *b* charging cost [€/kWh]	Battery	0
$P_{b,t}^{dis}$	Battery *b* discharging cost [€/kWh]	Battery	0.1
G_t^{sch}	PV panel expected production in period *t*, scenario dependent [kWh]	Generator	(0-22.8)
P_t^G	Compensation for reducing generation output of the PV panel [€/kWh]	Generator	0.05
W_t^{l-inf}	Inflexible load expected consumption in period *t*, scenario dependent [kWh]	Load	(0-14.57)

5.1.2. Design Variables

Symbol	Explanation	Component
δ_t^b	Binary variable. Takes a value of 1 if electricity is being bought from the grid in period *t*, otherwise takes a value of 0.	Grid
δ_t^s	Binary variable. Takes a value of 1 if electricity is being sold to the grid in period *t*, otherwise takes a value of 0.	Grid
χ_t^b	Continuous non-negative variable. Electricity bought in period *t* [kWh]	Grid
χ_t^s	Continuous non-negative variable. Electricity sold in period *t* [kWh]	Grid
z	Continuous variable. Total net cost for electricity exchanged with the grid in all periods of the planning horizon [€].	Grid
φ_t^g	Continuous non-negative variable. Electricity produced from PV panel in period *t* [kWh].	Generator
$\sigma_{b,t}^{ch}$	Continuous non-negative variable. Electricity charged to battery *b* in period *t* [kWh]	Battery
$\sigma_{b,t}^{dis}$	Continuous non-negative variable. Electricity discharged from battery *b* in period *t* [kWh]	Battery
$\sigma_{b,t}^{soc}$	Continuous non-negative variable. Electricity stored in battery *b* in period *t* [kWh].	Battery
ξ^{ch}	Continuous non-negative variable. Total cost of charging the set of batteries in all periods of the planning horizon [€]	Battery
ξ^{dis}	Continuous non-negative variable. Total cost of discharging the set of batteries in all periods of the planning horizon [€]	Battery
ξ^{gen}	Continuous non-negative variable. Total compensation for reducing PV generation in all periods of the planning horizon [€]	Generator
ω_t^l	Continuous non-negative variable. Electricity consumed by the inflexible load in period *t* [kWh]	Load

5.1.3. Design Functions

The equations in this section describe the different constraints that are established by the model in order to have a correct performance of the HEMS components and no physical laws are violated. Furthermore, the system reliability is kept.

5.1.3.1. Battery models

The battery SOC $\sigma_{b,t}^{soc}$ at period t is dependent on the SOC of the previous period $t-1$, the amount of energy charged $\sigma_{b,t}^{ch}$ and discharged $\sigma_{b,t}^{dis}$ in period t considering the corresponding efficiencies ($A_b^{ch}$and A_b^{dis}). These efficiencies are assumed as typical values. The SOC evolution constraint is modelled as follows:

$$\sigma_{b,t}^{soc} = \sigma_{b,t-1}^{soc} + A_b^{ch} * \sigma_{b,t}^{ch} - \frac{\sigma_{b,t}^{dis}}{A_b^{dis}},\ \forall\, b \in B, t \in\ T_{restr} \qquad \text{Equation 12}$$

The initial SOC $\sigma_{b,0}^{soc}$ is an input parameter, therefore the equation above is modelled using T_{restr}. The initial SOC is dependent on the parameter $\sigma_{b,0}^{soc-in}$, the amount of energy charged $\sigma_{b,0}^{ch}$ and discharged $\sigma_{b,0}^{dis}$ in $t=0$ considering the corresponding efficiencies ($A_b^{ch}$and A_b^{dis}). The initial SOC is modelled as follows.

$$\sigma_{b,0}^{soc} = \sigma_{b,0}^{soc-in} + A_b^{ch} * \sigma_{b,0}^{ch} - \frac{\sigma_{b,0}^{dis}}{A_b^{dis}},\ \forall\, b \in B, t \in\ 0 \qquad \text{Equation 13}$$

The SOC limits for avoiding deep discharge O_b^{min} or overcharge O_b^{max} were assumed as typical values from data sheets. The model is given by:

$$O_b^{min} \leq \sigma_{b,t}^{soc} \leq O_b^{max},\ \forall\, b \in B, t \in T \qquad \text{Equation 14}$$

The limit on the charging Q_b^{ch} and discharging Q_b^{dis} powers is also given by data sheets and the model is:

$$\sigma_{b,t}^{ch} \leq Q_b^{ch}, \; \forall \, b \in B, t \in T \qquad \text{Equation 15}$$

$$\sigma_{b,t}^{dis} \leq Q_b^{dis}, \; \forall \, b \in B, t \in T \qquad \text{Equation 16}$$

As mentioned before, the INVADE project shapes the charging and discharging profiles according to the SOC. There is a charging threshold S_b^{ch}, normally set at a typical value of 80% of the maximum SOC O_b^{max}, after which the charging power is linearly decreased until the battery reaches 100. Figure 20 shows the charging power profile and the equation is as follows:

$$\sigma_{b,t}^{ch} \leq \frac{-Q_b^{ch}}{1-S_b^{ch}}\left(\frac{\sigma_{b,t}^{soc}}{O_b^{max}} - 1\right) \qquad \text{Equation 17}$$

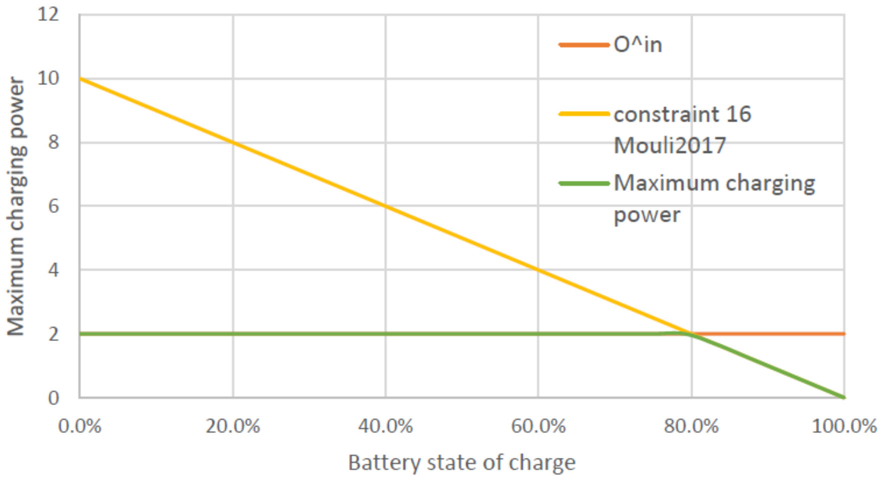

Figure 20. Charging profile constraint for the battery

Similarly, there is a discharging threshold S_b^{dis}, normally set at a typical value of 10% of the maximum SOC, O_b^{max}, below which the discharging power is linearly decreased until the minimum SOC is reached. Figure 21 shows the discharging power profile, and the equation is as follows:

$$\sigma_{b,t}^{dis} \leq \frac{Q_b^{dis}}{S_b^{dis}}\left(\frac{\sigma_{b,t}^{soc}}{O_b^{max}}\right) \qquad \text{Equation 18}$$

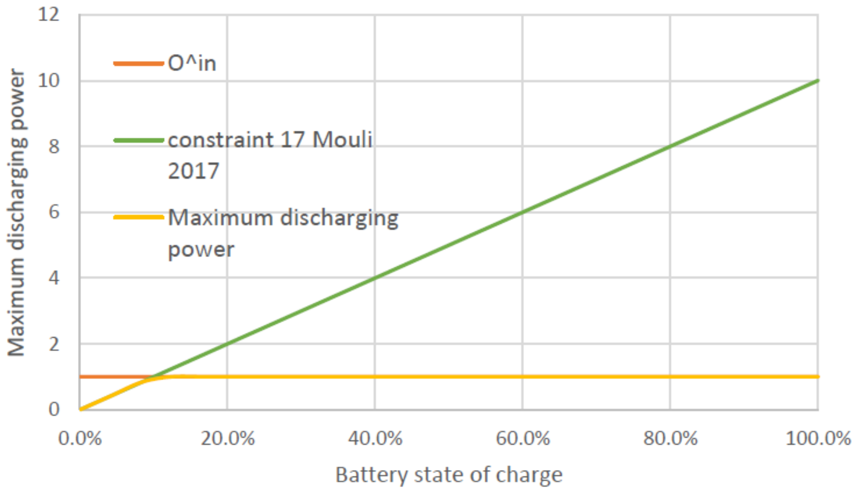

Figure 21. Discharging profile constraint for the battery

5.1.3.2. Load model

As mentioned, the load in this project is considered inflexible. In this case, the scheduled load $\omega_{l,t}$ must be equal to the predicted load $W_{l,t}^{load}$.

$$\omega_t^l = W_t^{l-inf} \forall \ t \in T \qquad \text{Equation 19}$$

5.1.3.3. Generator models

The generation unit models used in the project are curtailable reducible. Therefore, the scheduled generation output $\psi_{g,t}$ can range from 0 up to the predicted generation $W_{g,t}^{prod}$. The model used is:

$$0 \le \varphi_t^g \le G_t^{sch}, \ \forall \, t \in T \qquad \text{Equation 20}$$

5.1.3.4. Electricity balances

For each period in the planning horizon, power enters the HEMS either by the grid χ_t^b, the battery $\sum_{b\in B} \sigma_{b,t}^{dis}$ or is generated by the PV system φ_t^g. Power exits the system when it is supplied to the appliances ω_t^l, sold to the grid χ_t^s, or when the battery is charged $\sum_{b\in B} \sigma_{b,t}^{ch}$. This can be seen in

ETSEIB

the following equation.

$$\varphi_t^g + \sum_{b\in B} \sigma_{b,t}^{dis} + \chi_t^b = \chi_t^s + \omega_t^l + \sum_{b\in B} \sigma_{b,t}^{ch}, \qquad \forall\, t \in T \qquad \text{Equation 21}$$

The simultaneity constraint for buying and selling to the grid is given with binary variables for buying δ_t^b and selling δ_t^s by:

$$\delta_t^b + \delta_t^s \leq 1,\ \forall\, t \in T \qquad \text{Equation 22}$$

The grid importing and exporting limits, physical or economic, are represented by the parameters X^{imp} and X^{exp}, respectively are given by:

$$\chi_t^b \leq \delta_t^b X^{imp},\ \forall\, t \in T \qquad \text{Equation 23}$$
$$\chi_t^s \leq \delta_t^{sell} X^{exp},\ \forall\, t \in T \qquad \text{Equation 24}$$

As stated in Section 3, the power balances are highly dependent on the application. In case of the INVADE project, the Bulgarian pilot does not allow injection of energy into the grid which makes $X^{exp} = 0$ kW. In the Norwegian pilot the export capacity, is a regulatory limit which means $X^{exp} = 100$ kW.

5.1.4. Objective Function

For this thesis the objective function is the minimization of costs. However, since the costs occur in different stages it is necessary to divide them accordingly, to avoid anticipativity solutions. The total costs are the costs for charging and discharging the battery, the remuneration for curtailing generation and the costs of buying and selling energy from and to the grid, respectively.

As stated, this project is developed mainly in two stages. Therefore, the above-mentioned costs are divided in First Stage Costs and Second Stage Costs.

The objective function consists of the minimization of the First Stage Costs and the expected Second Stage Cost. Therefore, per scenario the total cost can be defined as the sum of both stages' costs.

First Stage Costs

These costs are related to the battery since the decisions regarding its operations are done in the First Stage.

Charging of the battery degrades it, therefore there is a cost associated to it $P^{ch}_{b,t}$, which is proportional to the energy charged $\sigma^{ch}_{b,t}$. This value was chosen as 0, because no consensus has been found. However, it could involve a more complex modelling and be pursued in further research. The total cost of the battery charging flexibility ζ^{ch} is given by:

$$\zeta^{ch} = \sum_{b\in B}\sum_{t\in T} P^{ch}_{b,t} * \sigma^{ch}_{b,t}\,, \ \forall\, b \in B, t \in T \qquad \text{Equation 25}$$

Similarly, discharging of the battery degrades it, therefore there is a cost associated to it $P^{dis}_{b,t}$ and it is proportional to the energy discharged $\sigma^{dis}_{b,t}$. This value was chosen as a standard value as research by other colleagues is being done. However, this research is no longer addressed in the present work. The total cost of the battery discharging flexibility ζ^{dis} is given by:

$$\zeta^{dis} = \sum_{b\in B}\sum_{t\in T} P^{dis}_{b,t} * \sigma^{dis}_{b,t}\,, \ \forall\, b \in B, t \in T \qquad \text{Equation 26}$$

The First Stage Costs is the sum of ζ^{ch} and ζ^{dis}.

Second Stage Costs

These costs consist of the electricity bought and sold to the grid as well as the compensation for curtailing generation. They are considered in the second stage since their values are dependent on the scenario output and the recourse actions, therefore the decisions with respect to these variables are done in the Second Stage.

In the Norwegian pilot as mentioned in Section 1.3, the focus is on the prosumers and the great value for flexibility offered by their tariffs. The structure of the Norwegian tariff can be found in [13]. It includes cost for energy, grid use and taxes. As well, energy and flexibility services can be sold to the grid.

For the prosumer, the main objective is to minimize costs. The total cost for energy exchanged with the grid z is calculated with:

$$z = \sum_{t\in T}(P_t^{r-b} + P_t^{g-b} + P_t^{tax})\chi_t^b P^{VAT} - (P_t^{r-s} + P_t^{g-s})\chi_t^s \qquad \text{Equation 27}$$

The objective minimizes costs for all the periods t within the planning horizon T. In the Norwegian pilot, a prosumer can have a retail contract with a price P_t^{r-b} for buying and $P_t^{r-s}$for selling. As well, it is possible to have a grid contract, with a price $P_t^{g-b}$for buying and P_t^{g-s} for selling. All four are proportional to the amount of energy bought χ_t^b and sold χ_t^s, respectively.

Furthermore, a P_t^{tax}, proportional to the energy bought χ_t^b , accounts for grid use and an energy fund [67]. The Value Added Tax (VAT) is calculated over the total cost for bought energy, and currently in Norway is of 25%. Therefore P^{VAT} = 1.25.

If the generation is reduced to give flexibility to the system and there is a payment for this service, the value of such flexibility ζ^{gen} is given by:

$$\zeta^{gen} = \sum_{t\in T} P_t^G\left(G_t^{sch} - \varphi_t^g\right), \ \forall \ t \in T \qquad \text{Equation 28}$$

For the curtailment of generation, in the Norwegian pilot, the end user decides the cost of curtailing generation P_t^G. It is estimated to range between 0.08 and 0.1 €/kWh. However, due to a low certainty a conservative approach was used and for this project P_t^G= 0.05 €/kWh.

5.2. Software

As stated before, an important objective of this work was to use open software in all stages. Most of the algorithm was carried out in Python, however R was used for the statistical analysis of the data.

As it can be seen in Figure 22, Python is becoming the most popular programming language worldwide due to its simplicity and flexibility [68]. These characteristics have proven a great advantage, specially while reading and understanding others' syntaxes. Furthermore, Python based codes are easy to build-on and it is easily translatable to other languages [69], making a great choice for big projects like the INVADE. The huge online community provides readily available support which encourages creativity and as well, there are a vast amount of supporting libraries [2] like the ones used in this thesis.

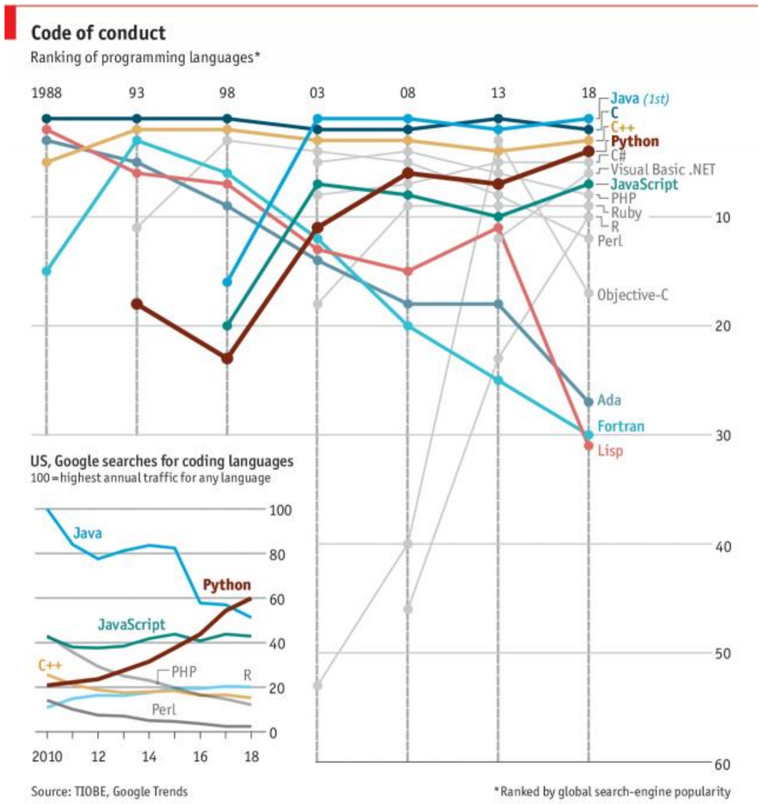

Figure 22. Python is becoming the favorite coding language [68]

On the other hand, R is an open source statistical programming environment and language. It is rapidly becoming the preferred tool for data analysis in a wide range of disciplines [70]. R was developed specifically for statistical analysis and is particularly powerful to explore datasets [71]. Even though, similar analysis can be carried out in Python, the visual representations, the decomposition of datasets, the easiness to manipulate and analyze them, as well as the statistics libraries available were quite useful to simplify the analysis and creation of scenarios for this thesis.

5.2.1. R

The R libraries used were:

Dplyr

Also called the Grammar of Data Manipulation, Dplyr allows to manipulate, filter and group datasets. Afterwards, the statistical analysis was easily performed [72].

Ggplot2

It Is a library based on the grammar of graphics. The main purpose is useful to present information in an organized and clear manner.

Xlsx

It is a package that allows R to read, write and format Excel files. This was specific to present the scenarios in a way that are easily read by the main python-based algorithm.

5.2.2. Python

The Python libraries used were:

Pandas

It is an open source library for managing data structures and data analysis [73]. As mentioned before, it was used to process and manage the input data and output data.

Json

This library allows the main program to open the results given by the optimization solver. After the optimization is done the results can be stored in a json or a csv file. However, the format of the csv made it difficult to parse and obtain valuable information automatically, therefore it was opted to use the json option.

Os

It is used to open the system terminal to call the *runph* PySP command for solving the optimization algorithm.

NetworkX

This is a package used "for the creation, manipulation and study of the structure, dynamics and functions of complex networks [74]". It was used to create the scenario tree structure from a Pandas data frame.

Pyomo

It stands for "Python Optimization Modelling Objects", and as its name states it, it is a python-based tool used for mathematical modelling [2]. Furthermore, as it can be seen in Figure 23, Pyomo is an open source and free optimization environment that covers the different stages of optimization: data input, data manipulation, optimization, analysis and visualization [75].

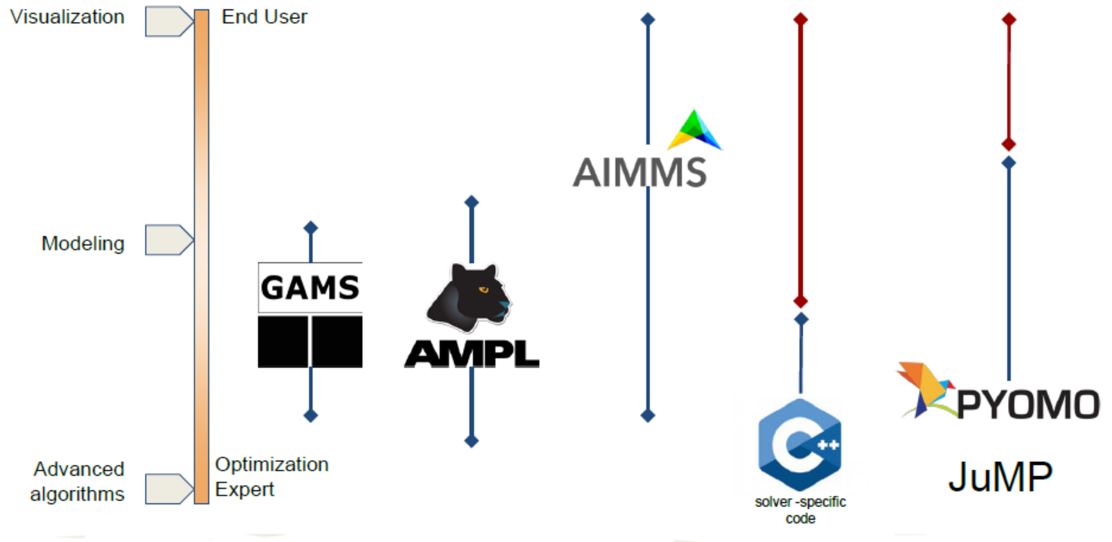

Figure 23. Optimization Environments Overview [75]

Pyomo is object-oriented, hence the requirements, stated before such as the quantitative model, design variables and design functions are all Pyomo objects with specific attributes. The optimization objects created in Pyomo are Sets, Parameters, Variables, Constraints, Objective, Model and Solver.

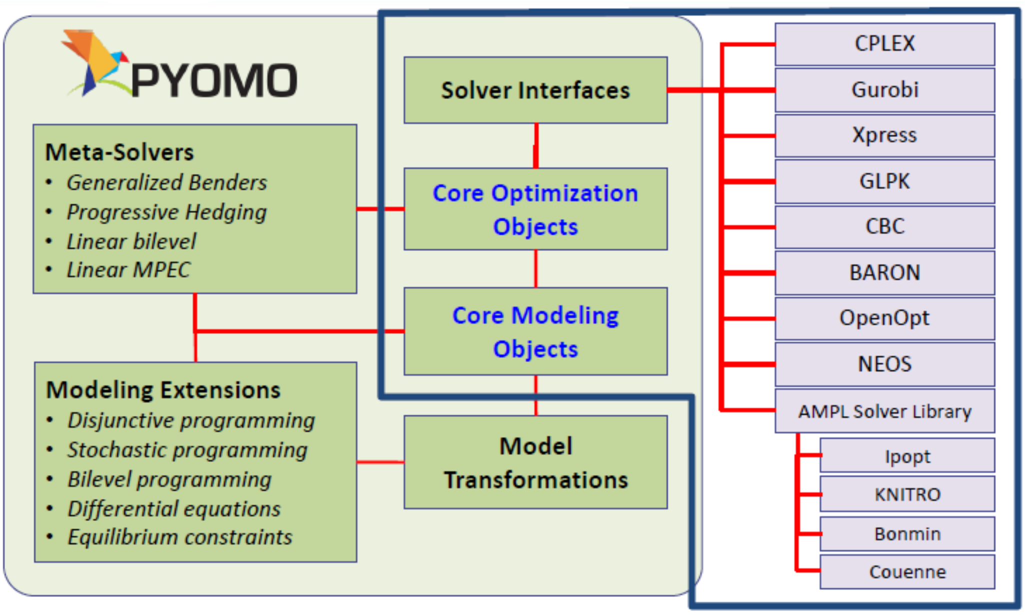

Figure 24. Pyomo Modelling Objects [75]

As Figure 24, shows, Pyomo has a modelling extension for SP called PySP, which was the one used in this work. Also, it can be appreciated the different meta-solvers and solver interfaces. For this project Progressive Hedging and Gurobi were used, respectively.

PySP

PySP is an extension of the Pyomo package designed to model objects for SP [2]. PySP allows to extend a deterministic model into a stochastic program and solve it. By being Python based, PySP is also characterized with simplicity, therefore it is a good option for non-specialists in stochastic optimization. In this way, only a shallow understanding of the solving algorithms is required [2].

There are other ways to do simple stochastic optimization with Pyomo and Python, however, as it is specifically developed to solve SP more complex problems can be solved and it is simpler to present the data. Moreover, as PySP is an extension of Pyomo, it has full access to Pyomo's solvers and models.

PySP supports two commands to solve stochastic problems: the *runef* command for the Extensive Form and the *runph* command for the PHA. Another advantage of PySP is that the solver is highly configurable [2]. For example, the Watson-Woodruff extensions accelerate the algorithm to converge in less iterations. Within Pyomo the Benders meta-solver can be used, however, this was not explored in this work.

The *runph* meta-solver needs to have an initial value for iterating the unknown variables. This is set using the default-rho command. It is important to choose an appropriate value to avoid increasing the time of convergence. Furthermore, since the PHA uses squared penalty terms, solvers like GLPK require that these terms are linearized. On the other hand, the solver used for this thesis, Gurobi, can handle the penalty terms.

5.3. Input Data

As mentioned before the data is mainly presented as excel files X1-X3 and as a csv file C1 and is imported by R1and P2.

5.3.1. Stochastic Data

The example data used for the analysis was downloaded from [76] using the ID 5357, and consisted of PV generation and the energy used by the loads. The grid consumption could be calculated with $Load = Generation + Grid$ [77]. From this, it was inferred that the house did not possess a battery or other generator. The data consisted of hourly values for the year 2016. The units of the parameters are kWh. The data was stored in C1 which was imported by R1.

Afterwards, using Dplyr, the data was separated into two data frames, generation and consumption. First each data frame was grouped by month and the probability distribution analyzed.

With these, the month of March was chosen as it had an average behavior and dispersion for both parameters. For the generation, negative values were taken as 0 as they can be attributed to measurement error.

The data belonging to March was grouped by hour. For the generation. only the hours between 8 and 18 hours were chosen. This was to reduce the impact of the night and low generation of the first and last hours of the day that distort the probability distribution: The values would fall into a one tail distribution instead of a normal distribution. Afterwards, the per hour average (μ) and standard deviation (σ) was obtained for both parameters.

For the creation of scenarios, R1 can generate as many scenarios as required. In the present analysis, 10 scenarios were used. With the average and standard distribution per hour, a normal distribution was assumed and divided in deciles, each with a 10% probability. Then, each scenario was constructed by adding the corresponding decile per hour.

As mentioned before, the creation of scenarios is not the focus of this work and several assumptions were made. First the hourly values are independent from each other. Secondly, a normal distribution per hour is assumed, in the data given this is true for the hours 12,13,14, however the first and last hours follow more of a one tailed distribution. For the load, it was assumed that the day of the week has no influence on the consumption.

These assumptions create simple scenarios that might not hold true for every case, and a deeper analysis should be done with respect to the pilot projects. However, the lack of data made it difficult to create valuable scenarios. A next step for this research would be to use K-means clustering as proposed by [19] with more data This could be done with R.

To study the effect of the previous assumptions, a sensitivity analysis of the input data was done by creating six sets of scenarios for each stochastic parameter. The hourly σ previously obtained was scaled up and scaled down to represent possible cases of input data. In total there were six sets of ten different scenarios, all with the same hourly mean but with different hourly σ. The scaling factors used were: 0.5 σ, 0.707 σ, 1 σ, 1.5 σ, 2 σ and 2.24 σ.

The values were rounded to four decimal places, the data frames formatted as to meet the requirements of P2 and then saved into X4. This file has two sheets, one per stochastic parameter (generation and load), the index of each sheet is the hour and the columns names are the scenarios.

5.3.2. Deterministic Data

The deterministic data is given in X1 and X2. X1 contains four sheets: prosumer capacity, electricity prices, the VAT, and the curtailing generator cost. The information about the electricity prices is indexed by hours and the columns refer to the retail and grid prices for buying and selling energy to the grid. The prosumer capacity is indexed by the ID of the prosumer.

X2 contains the parameters of the batteries, it consists of only one sheet which is indexed by the battery ID. The columns refer to all the parameters of batteries required for the model. The battery was sized accordingly to the data obtained from [76].

5.3.3. Scenario Tree Data

The scenario tree data is given in X3. The file contains three sheets and follows the requirements needed by NetworkX to create the scenario tree. The first sheet describes the nodes of the tree. Each children node is associated to a father with a probability. For simplicity, the last stage nodes were directly named with the corresponding scenario name. This work had ten children nodes (Scenario1-Scenario10) and one father node (Root Node). Each of the children is connected to the Root Node with a 10% probability. The index of this sheet is the children node names, which in this case is the scenario names. Caution should be taken when creating a multistage tree.

The second sheet states the number of stages and the costs associated to each of them. For this work, there are two stages each with the First Stage Cost and Second Stage Cost which compose the objective function.

Finally, the third sheet assigns the design variables to the corresponding decision-making stage. As stated, in this thesis there are nine design variables. The battery SOC, charging and discharging variables are set in the first stage, therefore they are allocated to this stage. On the contrary, the generation, load and the buying/selling to the grid variables are assigned to the second stage because its value depends on the information disclosed in the second stage.

X3 is read by pandas in P2. The scenario names and index are passed to P4 to incorporate to the reference model. The scenario tree object is created in P3 using the libraries NetworkX and Pyomo and the data provided by X3.

5.4. Stochastic Model

P4 consists of the creation of the model object, it is based on the mathematical model developed by INVADE and initializes the object using the data provided by P2.

The stochastic model created, based on the mathematical model presented before was done in Pyomo. The type of model constructed is a concrete model. In this type of models, the values for parameters and sets must be stated and initialized in the moment of the object creation. A concrete model was chosen, against an abstract model, to support the automation of the scenario tree creation. The built- in functions for instance creation and scenario tree creation only support concrete models.

An instance is created with the function *pysp_instance_creation_callback* which takes the scenario name and the node names as inputs. After the model is created, and the deterministic parameters and sets initialized, the scenario-dependent parameters are stored in a dictionary with the key being the scenario name. The model is cloned and the values per scenario of the stochastic parameters assigned to each cloned instance.

Every time P4 is called it generates an instance. This instance consists of the concrete model previously described and the inputs are scenario name and the different nodes involved in that scenario. When creating the concrete model, as stated before the data was imported from P2. For the stochastic variables, a dictionary is given as input with the key being the scenario name. Since the input for the instance callback is the scenario name the stochastic parameter is initialized from the dictionary using the respective scenario.

The model is then cloned to create different instances for different scenarios and the instance is then return to P1.

5.5. Main Program

P1 calls the Pyomo environment functions, the Pandas, Os and Json libraries. As well, it directly imports the values of the parameters given by P2.

The PySP extension is executed from the system terminal. As mentioned before, the progressive hedging meta-solver was used for the solving of this work. The command to call it is *runph*. Then it is required to specify the name of the file where the Reference Model is, the solver to use for the optimization, and the default rho value. As well, the solution writer command was used to store the results in a file.

The objects required by the PySP extension are the instances created by P3, the scenario tree created by P2 and the solver specified within P1-which in case of this thesis is Gurobi. The solver then uses the PHA to reach a solution.

As previously stated, after the optimization is done the results are stored in a json file. P1 extracts the results from such file and creates relevant outputs stored in X5. The output file X5 has five sheets. The first shows the schedule of the batteries SOC, charging and discharging and it is indexed by hour. The second sheet has the generation indexed by hour. For the stochastic optimization the columns give the different scenarios' generation, for the robust and deterministic cases there is only one column. The third sheet following the same structure as the second one, states the load consumption. The fourth, shows the buying and selling of energy to the grid indexed by hour, scenario dependent. Finally, the fifth sheet, shows the result of the optimization which is the cost the prosumer would pay in the end of the day. For the stochastic optimization the cost is divided in stages and then the expected cost is calculated. For the robust optimization the cost shows the worst-case-scenario cost, whereas the deterministic optimization shows the expected or average cost.

6. Results

As stated in Section 2.1, the main objective of this work was to create a stochastic optimization for a HEMS within the INVADE project. To assess the performance of the algorithm developed, three case studies with six subcases each were created. Additionally, the deterministic and robust benchmark cases, DB and RB respectively, were also studied. In total two hundred simulations were run in this thesis and the results are presented in this section.

The three case studies are:

- CS1 = Generation and consumption are stochastic parameters.
- CS2 = Generation is a stochastic parameter and consumption a deterministic one.
- CS3 = Consumption is a stochastic parameter and generation a deterministic one.

In Section 5.3.1, the limitations on the scenarios were described. As to compensate for the lack of valuable example data, a sensitivity analysis was performed. This analysis consisted of the creation of six subcases per case study. Each subcase had a different set of ten scenarios. All the subcases had the same mean value, but the hourly σ of each stochastic parameter considered was scaled by factors 0.5, 0.707, 1, 1.5, 2 and 2.24, respectively.

For the deterministic parameters of CS2 and CS3 as well as for DB, the mean or expected values were used. For RB the worst-case-scenario data of CS1: σ1 was used, this corresponds to the scenario 10 data.

6.1. Comparison between the stochastic and the deterministic approach

Overall, the stochastic optimization and the deterministic optimization solutions are equivalent, with the stochastic optimization having a slight trend of improvement when the data has a higher σ. The main reason is that the elasticity of the grid absorbs the uncertainty, thus there is no need to mitigate the loss of profit. A further step for research would be the incorporation of several prosumers and adding more constraints to the grid, as to observe whether this tendency holds.

As it can be seen in Figure 25, the buying and selling profiles are essentially the same for CS1:

σ1, CS2: σ1, CS3: σ1 and DB. The RB, naturally, has a higher consumption in both evening and afternoon peaks as a consequence of the input data which has a higher load. It is interesting to see, that RB is the only simulation that sells electricity to the grid. This could be considered as the last resort of the HEMS to minimize costs and therefore it is not as attractive as other flexibility options.

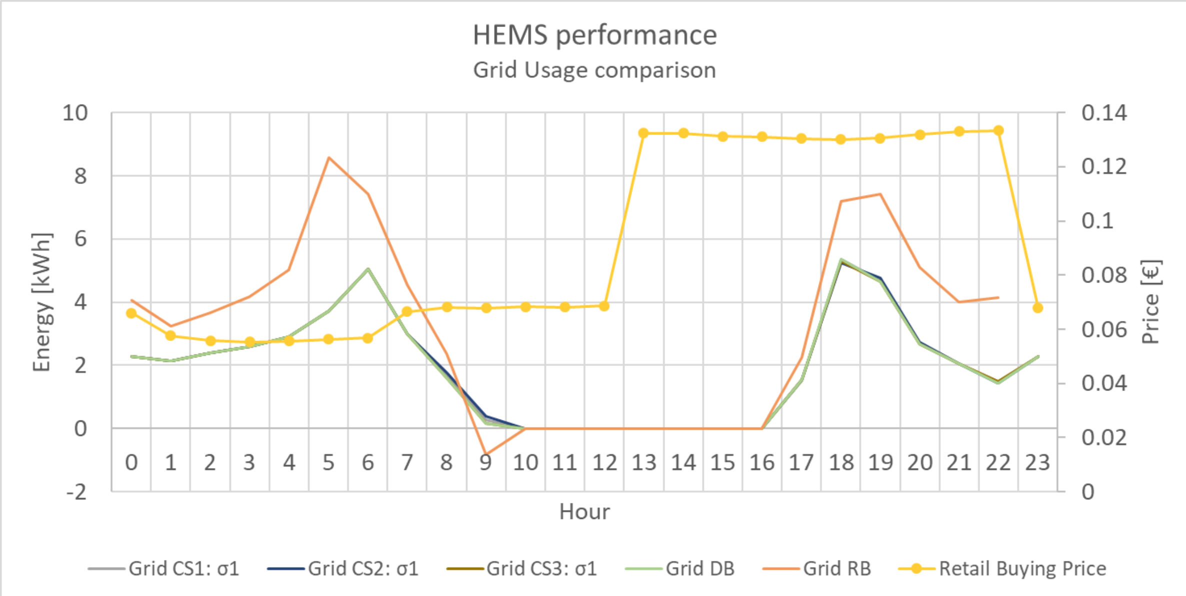

Figure 25. Grid usage comparison among the three case studies and the two benchmarks

Figure 26 shows the scheduling of the SOC for CS1: σ1, CS2: σ1, CS3: σ1, DB and RB. CS2: σ1, copes with the uncertainty of the generation by drawing more energy from the battery. RB follows a cautionary approach and charges the battery as fast as possible to cover the most of the high price periods. CS3: σ1, CS3: σ1 and DB have a quite similar profile.

The stochasticity of the generation has a higher impact on the SOC, whereas the stochasticity loads offsets this effect. This can be explained by the different profiles of the scenario data. The load has a smaller variability than the generation and there is an opposite nature between the two parameters- consumption and generation.

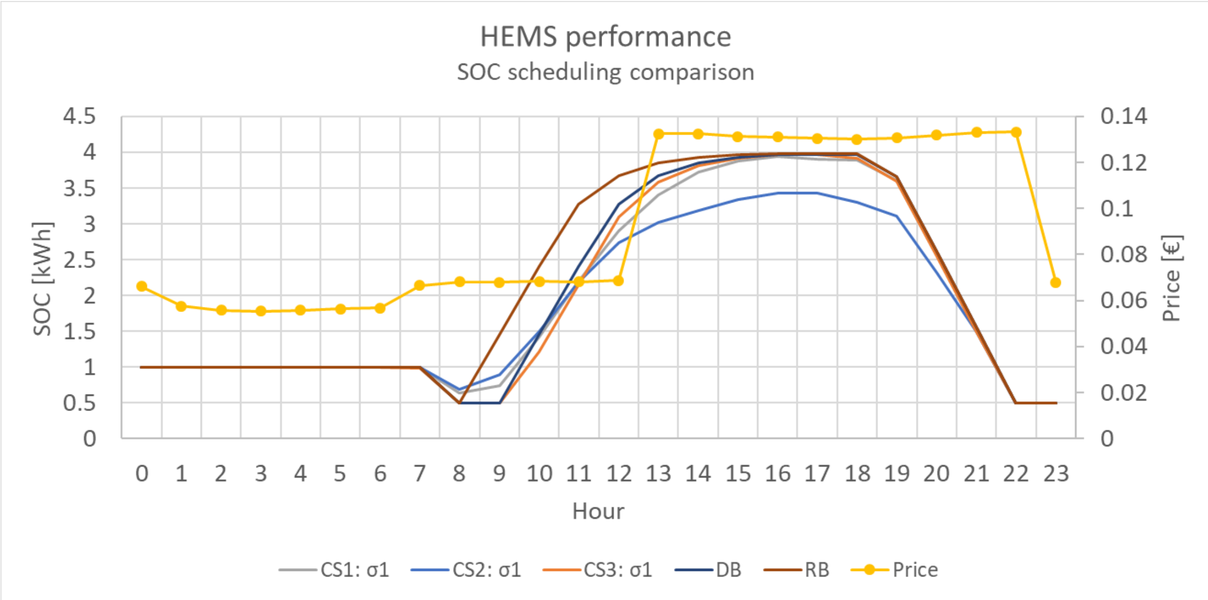

Figure 26. SOC comparison among the three case studies and the two benchmarks

As it can be seen in Figure 27, generation curtailment is done in all of the simulations. The curtailment profiles are fairly similar between CS1: σ1, CS2: σ1, CS3: σ1 and DB. RB follows a curtailment even with a generation significantly smaller. Consequently, it can be stated that the curtailment price is a good signal for flexibility.

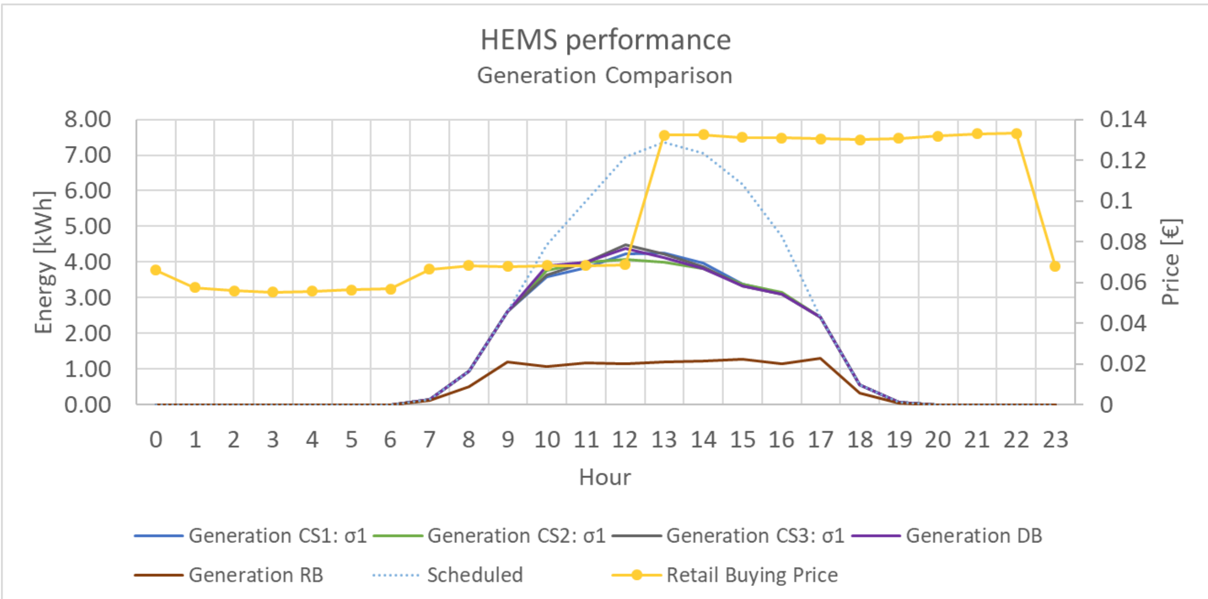

Figure 27. Generation curtailment comparison among the three case studies and the two benchmarks

After CS1: σ1, CS2: σ1 and CS3: σ1 were benchmarked with DB and RB the sensitivity analysis regarding the σ of the input data was done. As mentioned before, per case study, six sets of scenarios were created. For each set, the σ was scaled by a different factor. As an example, the distribution of the data for the period t=8 is shown in Figure 28. Similar patterns are followed each hour. From these graphs, is appreciated that for the period shown the consumption has an expected value with a high probability in σ0.5, σ0.707 and σ1, and even in the rest of the cases the expected value almost doubles the tail values. On the other hand, the generation forecast is spread out away from the mean, and in σ1.5, σ2 and σ2.24, there is no clear probability on an expected value.

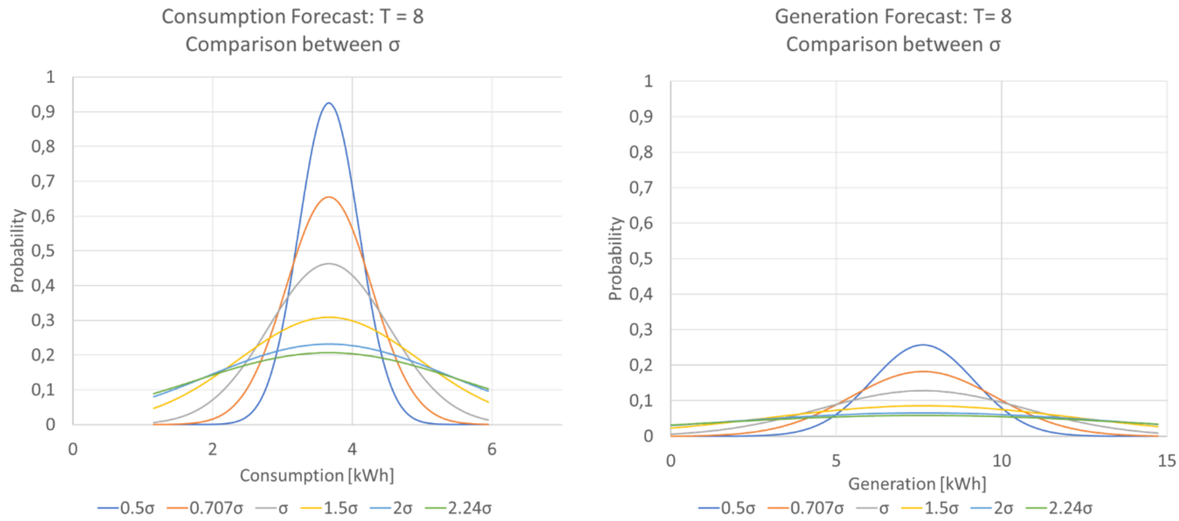

Figure 28. Example of generation forecasts and consumption forecasts used to create the subcases

The impact of the σ on the performance of the algorithm can be observed in Figure 29. As expected, the biggest differences in SOC are in the subcases with upscaling σ (factors 1.5, 2 and 2.24). CS1 has a difference of 1 kWh, between RD and σ2.24; for CS2, this difference is of 1.25 kWh and for CS3 the difference is of only 0.6 kWh.

For example, in CS2, there is a clear difference between downscale σ's (factors 0.5 and 0.707) and upscale factors. For σ0.5 and σ.707, the algorithm coincides exactly with the DB. The RB varies slightly by charging the battery earlier in the day. However, σ1, σ1.5, σ2 and σ2.24, differ in the scheduling of the SOC for the hours with high values of generation. The optimization algorithm tries to mitigate the stochastic nature of the generation and sets the SOC to a lower level, so the consumption will be met and avoid buying electricity from the grid.

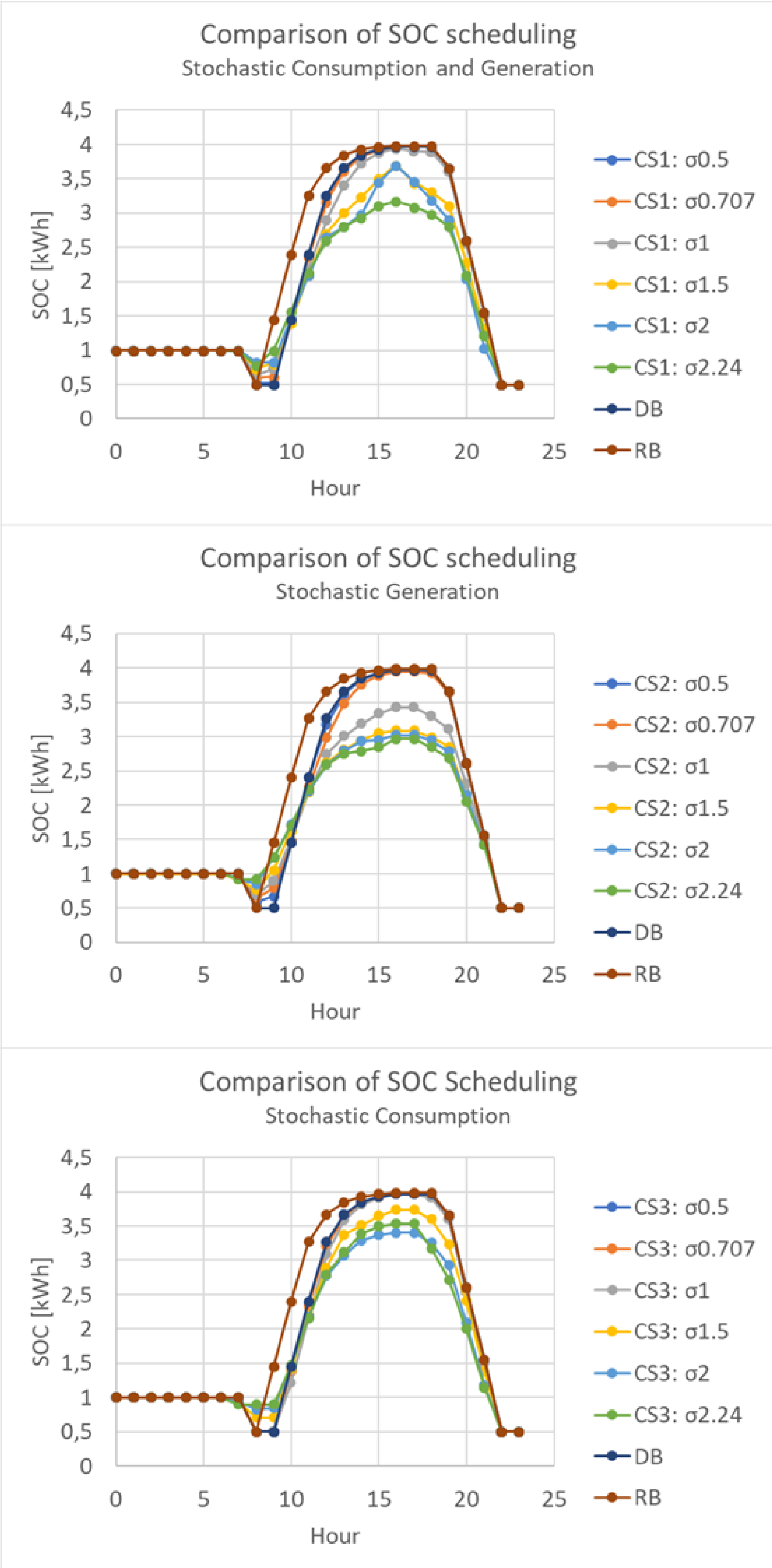

Figure 29. Effect of σ in SOC of the three case studies

The sensitivity analysis was done using the expected values for the different case studies and subcases. Due to the subcases having the same mean, they have the same expected value as well. Therefore, for the inflexible load, generation curtailment and the grid profile, the impact of σ was relatively small.

Regarding the effect of σ on generation curtailment, CS1, CS2 and CS3 have similar behaviors. There is curtailment in the peak generation hours, from 10 to 17 h. However, the differences per hour between the different subcases are relatively small.

The sensitivity analysis was also applied to the VSS from the different cases and subcases. In Figure 30, the CS1 costs for each scenario and σ is shown. Scenario 1 has the lowest cost and Scenario 10 has the highest cost for each subcase and case study. As it can be expected, the difference of costs between scenarios increases with increasing σ.

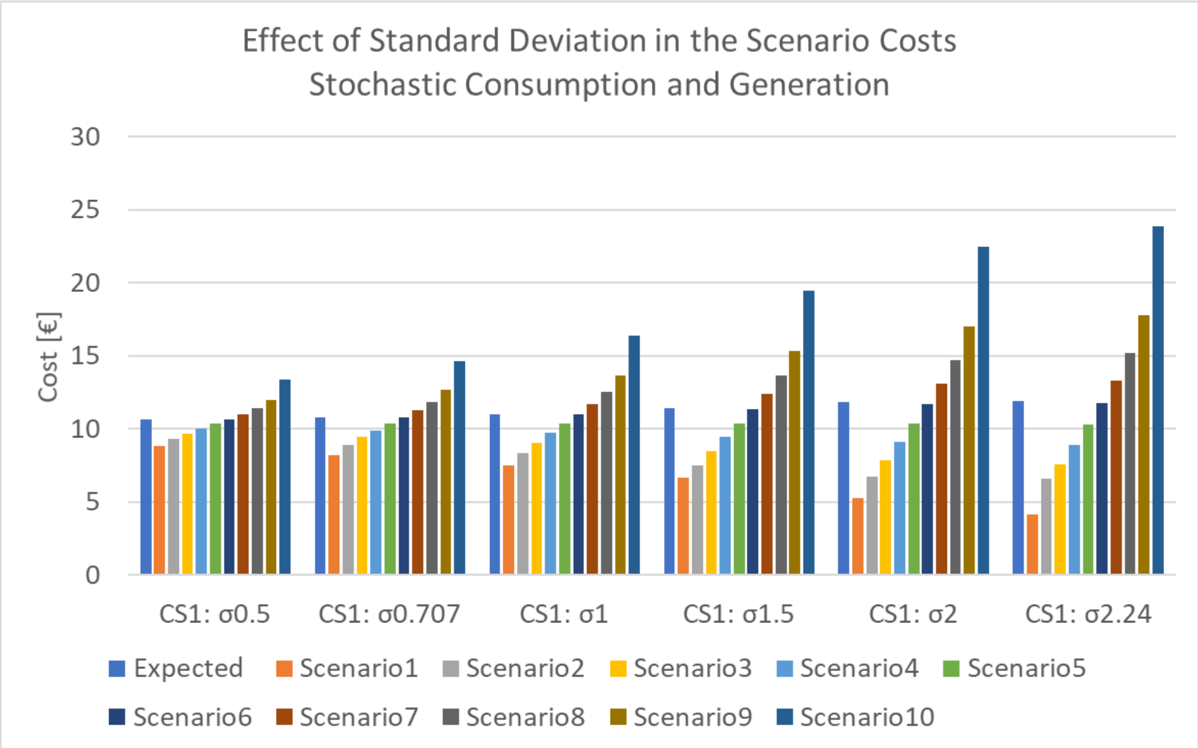

Figure 30. Example of impact of σ on scenario costs

These costs are complicated to compare, therefore the VSS was calcualteds to cuantify the value of stochastic optiimzation over the deterministic optimization. Furthermore, the sensitivy analysis can also be performed to the VSS.

For calculating the VSS of a subcase, the solution regarding the SOC obtained with DB were fixed as parameters. Then, deterministic optimizations were run for each of the scenarios and the costs obtained were weighted and summed. This value was then substracted from the expected value of the stochastic optimization to obtain the subcase VSS. This process was repeated for each subcase and case study, giving a total of 180 deterministic simulations. As it can be appreciated in Figure 31, the VSS is a negative number. This is due to the optimization problem being a minimization problem.

CS1 has a VSS of 0.002, CS2 of -0.08and CS3 of -0.042. Therefore, the stochastic optimization yields better results in CS2 and worst results in CS1.

For the different subcases a clear trend can be appreciated in Figure 31. The impact of σ follows an “s” behavior. With small values of σ, the VSS is close to zero. Then as σ increases, the VSS decreases abruptly, after a certain threshold, the VSS stops the abrupt minimization and follows a more constant decrease.

For CS1, the stochastic optimization becomes valuable only if the data has a σ bigger than 1. For CS2, the VSS decreases abruptly since 0.707 σ. Finally, for CS3 the VSS is always negative, however it increases steadily with σ bigger than 1.

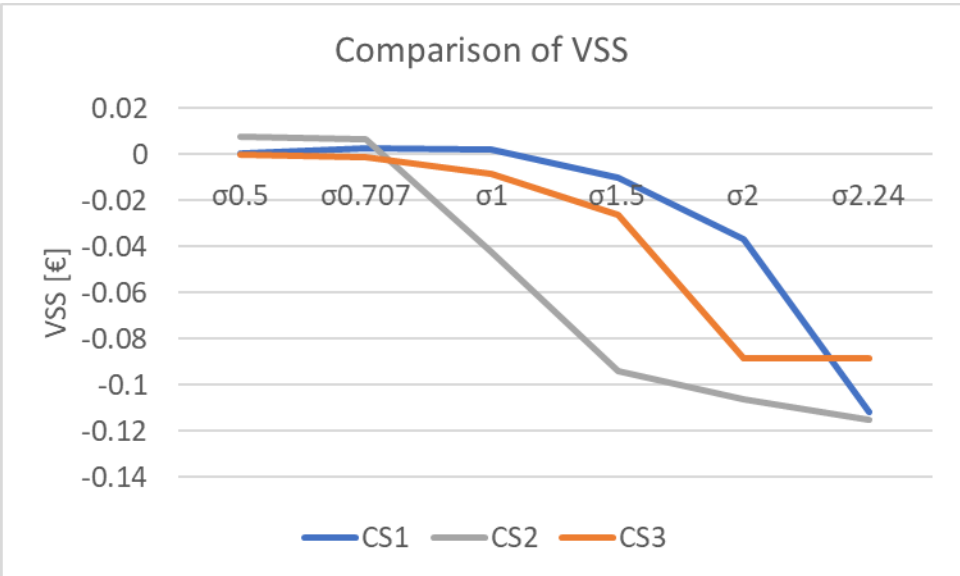

Figure 31. VSS of the three case studies

6.2. Quantification of flexibility

Figure 32 shows the HEMS performance during the day for CS1. The data shown for generation and consumption is the expected data. The main trend is that the algorithm schedules the components, so the load uses the PV production during the day. The morning consumption peak is covered entirely by the grid due to the low initial SOC, whereas the evening peak is only slightly shaved by using the battery. The main strategy of the algorithm is to reduce the consumption in the evening, where the price is constantly high.

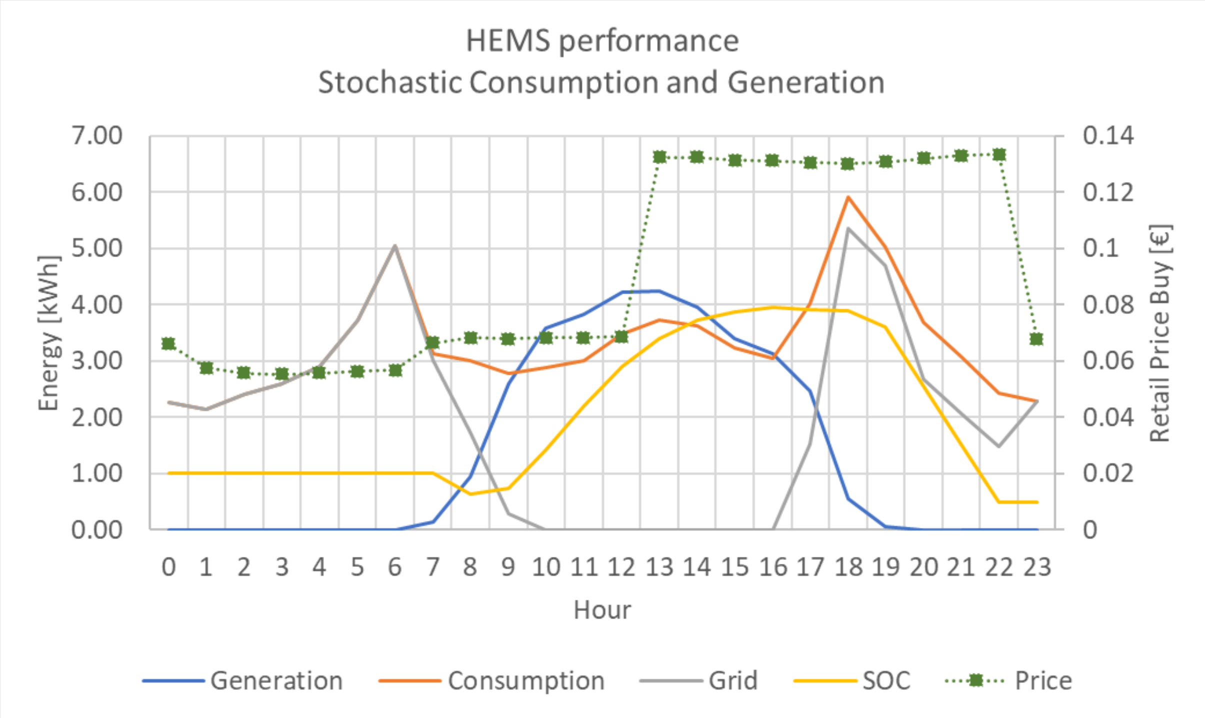

Figure 32. Scheduling of the HEMS for CS1

It can be appreciated that the price signals have a strong influence in the HEMS performance. In the morning when prices are low, there is no interest in shaving the peak. On the contrary, the evening peak is being reduced in response to the high costs, until the battery is discharged to its minimum. For all the case studies and the different subcases, the curtailment of PV starts at T=10 and finishes at T=17. This coincides with the periods time without grid consumption. A clear influence of the price signals is observed, as the highest costs of the evening are mitigated in all algorithms and the morning peak is not mitigated due to low prices. The total energy curtailed is similar in all the cases and subcases analyzed, with only small variations of no more than 0.5 kWh. The total average flexibility obtained by curtailing generation was obtained with Equation 28 is of:

$$\zeta^{gen} = 15.83 * \; P_t^G = 0.079 \; €$$

For the use of the battery, the SOC has differences according to the cases and subcases. The average flexibility offered by the battery was calculated by adding equation 25 and equation 26.

$$\zeta^{ch} + \zeta^{dis} = 3.5 * \; P_{b,t}^{ch} + 2.62 * P_{b,t}^{dis} = 0.262 \; €$$

The average flexibility offered by the stochastic optimization over the deterministic optimization was calculated with the absolute average VSS.

$$VSS = 0.04 \; €$$

In summary, the expected flexibility obtained with the stochastic algorithm is of:

$$Total\ Flexibility = \; \zeta^{gen} + \zeta^{ch} + \zeta^{dis} + VSS = 0.381 \; €$$

6.3. Use of open software

The optimization algorithm of this thesis was created using R and Python. Both confirmed why they are becoming the favorite programming languages: their simplicity and power.

R was useful to create scenarios, the easiness to read and write files was helpful when parsing and manipulating data, particularly the Dplyr library gives the opportunity filter and select data frames with whatever criteria desired and then perform statistical analysis on the selected data. The creation of scenarios was fast and automatic and thus, made the sensitivity analysis possible. The main disadvantage found is the syntax. Although it is fairly simple to use, time is needed to get acquainted with the different functions and language. As well, Rstudio has problems running in environments like Anaconda, which diminish its availability. It is also recommended for the creation of data visualization using the library ggplot2.

For the Python libraries, Pandas was a clear advantageous tool. This library allowed the connection between the different types of files. The simplicity of the data frame creation gives great power, when pursuing a real-life application and the data needs to be constantly updated.

Specifically, Pyomo allowed the translation of the mathematical model and the execution of the optimization in a simple way. The different objects for modelling recognized by Pyomo, makes it easy to focus on the performance of the algorithm rather than in the nuances of modelling. However, the PySP extension proved less helpful in translating the stochastic part of the model.

Although PySP might have great power for implementing SP, the lack of literature, examples and applications makes it hard to create even the simplest model. Much time was invested in looking for the correct functions to circumvent the standard format cited in literature, which is quite inflexible. This was finally managed through Pandas for the instance creation and NetworkX for the scenario tree creation. Most of the available PySP literature, as well as GitHub examples, are based on scenario data which is either randomly generated or needs to be manually updated.

The library NetworkX was used to replace a quite verbose *ScenarioTreeStructre.dat* file with a small python script. This automatized the scenario tree creation, which opened the possibility of increasing complexity and dynamism. Future research would use the algorithm to create multistage or dynamic scenario trees. Similarly, Pandas was used to automatize the instance creation, and this reduce repetition of information.

As soon as there were two stochastic parameters in the model, the extensive form was not enough to formulate the problem, therefore, the PH algorithm was needed. This is available in PySP using the *runph* command. This command, together with the solver power of Gurobi made the execution of the simulations fast and straightforward.

Finally, the performance of the Gurobi solver was satisfactory in all cases. The main advantage of Gurobi on other open software solver such as GLPK, is that it can handle non-linear terms. In the case of the PHA, the impossibility of handling non-linear terms, reduces the power of the algorithm.

Figure 33 shows the time required for doing the 18 stochastic simulations and the two deterministic simulations. As it can be appreciated there is no clear relationship between both variables. However, the iterative method of PH takes more time than the deterministic in most of the cases. Within the three cases to study, stochastic consumption, stochastic generation or both stochastic, there is no conclusion that can be drawn.

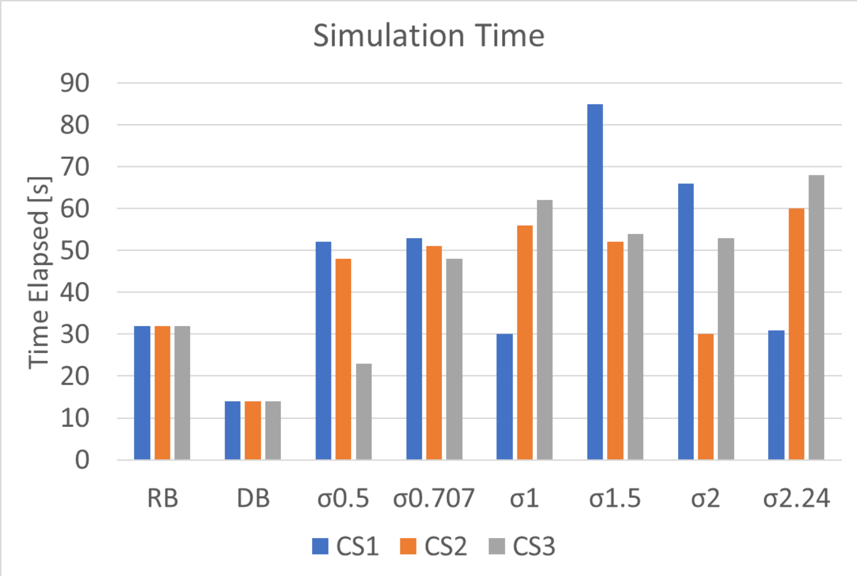

Figure 33. Time required for the three case studies' simulation

6.4. Environmental impact of the project

As stated in1.2, there is a need of cost-efficient solutions for mitigation of emissions. HEMS aim to do so, because any mitigation is done only by proper operation of the house components. However, it was found that the quantification of emission mitigation is complicated since they are related to the amount of energy not spent, which is a hypothetical quantity. Furthermore, externalities such as the use of servers and communication infrastructure are highly uncertain and can reduce the emissions mitigated This analysis is merely an overview of the impacts of this project, for a full Life Cycle Analysis on HEMS refer to: [78], [79].

In Norway, where the first pilot of the INVADE project will be implemented, the electricity carbon intensity is of 9 gCO_{2eq}/kWh [80]. Considering the average VSS of - 0.04 €, 114 Watts per day are being saved by using stochastic optimization. This accounts to 1.03 gCO_{2eq}/per day or 376 gCO_{2eq}/per year. This quantity should be scaled up, accordingly to the number of households in the pilot project.

The realization of this project, as well as the INVADE project, entails an environmental impact during its development. Specifically, regarding this thesis, the use of electronic devices and transport had environmental impacts of water, energy and residues. The approximation of emissions due to energy consumption is as shown in Table 2. The carbon intensity of consumption at low voltage in Spain is of 341 gCO_{2eq}/kWh [80].

Table 2. Electricity consumption emission quantification

Electric devices	Power [kW]	Hours of operation [h]	Energy consumption [kWh]	Emissions [$kgCO_{2eq}$]
Laptop	0.07	1800	126	42.97
External monitor	0.05	1300	65	22.17
HVAC system	0.125	400	50	17.05
Lights	0.02	2000	40	13.64
Router	0.0006	3600	2.16	0.74
Total				96.56

For the transportation, two means of transport were used. The bus which has a carbon intensity of 0.025 $kgCO_{2eq}$/ (km passenger). The subways which has a carbon intensity of 0.015 $kgCO_{2eq}$/(km passenger) [81]. The approximation of emissions due to transport usage is as it can be seen in Table 3.

Table 3. Public transport use emission quantification

Transport means	Distance[km]	Days [n]	Total distance [km]	Emissions [$kgCO_{2eq}$]
Subway	13	60	780	11.7
Bus	4	80	320	8
Total				19.7

The total emissions for this project was approximately 115.3 $kgCO_{2eq}$.

Conclusions

The goal of this work was to do an introductory assessment of SP, as an alternative to cope with uncertainty of consumption and generation in HEMS as part of the INVADE project. The problem presented is a MILP and the HEMS model created consists of curtailable PV generation, a battery, an inflexible load and the grid were enough to give an introduction on the stochastic modelling and fulfill the specific objectives

For the three case studies analyzed in this work, the stochastic optimization is equivalent to the deterministic one. The added value of the stochastic solution has an average of 0.04 € per day. CS2 performed better and CS1 worse, therefore, the stochasticity optimization performs better for coping with uncertainty of the generation than load uncertainty.

Moreover, the sensitivity analysis confirmed, that the value for the stochastic solution tends to increase in proportion to the scenarios' standard deviation. The strongest increase was observed in CS2. The grid's high flexibility absorbs most of the uncertainty which results in the same profiles for the different case studies, which explains the low values of VSS.

Regarding the HEMS performance, the algorithm proved effective in scheduling the components reacting to the market price signals, the low morning prices encourage consumption, and the high evening prices encourage peak shaving. The generation curtailment is attractive in all the case studies, whereas selling energy to the grid is only used in the worst-case scenario as the safe option. The battery was fully used in reactance to the price signals, confirming its usefulness as an important flexibility source. The expected flexibility was of 0.381 €.

With respect to the software used in this thesis, R allowed to simplify and automatize the data analysis and scenario generation. Pyomo excels in translating mathematical models in a simple manner. However, the extension for SP, PySP, is still lacking this advantage. Furthermore, the novelty of the extension, derives in a deficit of literature or examples.

The inflexible nature of the algorithm was circumvent using a combination of NetworkX and Pandas which allow for automatization and dynamism. Nevertheless, PySP and the PHA meta-solver available, allows for an increase in complexity of the problems studied. Furthermore, the commands and extensions within PySP increase its power. The solver preferred was Gurobi due to its capability of handling non-linear terms.

Finally, even though HEMS are considered as a cost-efficient strategy to mitigate GHG emissions, the potential and quantification of savings is difficult to do. For this thesis, stochastic optimization saves 114 Watts per day, which roughly translates to 1 gCO_{2eq}/per day. On the other hand, the emissions generated by the realization of this thesis were of 115 $kgCO_{2eq.}$

Stochastic optimization is increasing its popularity in the energy sector. Even though the case studies presented have a small VSS, further analysis should be conducted to find new areas of opportunity for this tool.

Acknowledgements

My sincere gratitude to the EIT for the opportunity of being part of this two-year journey.

The greatest thanks go to my parents who made me who I am. To my dad, my hero, my best teacher, who never lets me quit. To my mom, my best friend, my accomplice and the biggest fighter I know.

Thanks to Cristy for many years of love and care and to Pelusa, because you make our family whole.

Thanks to Pol Olivella and Andreas Sumper for the patience and attention To Maria Marin for all the help. Thanks to Katya Romo, Héctor Sarmiento, Ricardo Ganem and Juan Manuel Reyna, your knowledge has taken me this far.

Special thanks to Beto, Mayer and Luna who stuck up with me through these years and are the best friends one could ask for.

Thanks to Pia, Bryan and Sergio for making Barcelona an even nicer city. Thanks specially to Leon who motivated me through this whole project with laughter and endless cups of coffee.

Thanks to Laura and Renate for being the best friends through time and space.

Thanks to Anai, your guidance helps me grow as I never thought possible.

To Giulia, Srikant and Laura, my Leuven grandparents, I would not have made it without you!

Bibliography

[1] M. Beaudin and H. Zareipour, “Home energy management systems : A review of modelling and complexity,” vol. 45, pp. 318–335, 2015.

[2] W. E. Hart *et al.*, *Pyomo-optimization modeling in python*, vol. 67. Springer, 2012.

[3] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Second. .

[4] Merriam-Webster, “Standard Deviation.” 2018.

[5] Merriam-Webster, “Variance.” 2018.

[6] European Commission, “2020 Energy Strategy,” *Energy Strategy and Energy Union*, 2018. [Online]. Available: https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2020-energy-strategy. [Accessed: 18-Jul-2018].

[7] IPCC, *Summary for Policymakers*. 2014.

[8] Y. Mulugetta, E. Hertwich, K. Riahi, T. Gibon, and K. Neuhoff, “Climate Change 2014: Mitigation of climate change,” *IPCC Fifth Assess. Rep.*, pp. 527–532, 2014.

[9] International Energy Agency, “World Energy Balance 2010,” 2012. .

[10] International Energy Agency, “Share of Total Final Consumption by sector for World,” *Statistics*, 2017. [Online]. Available: https://www.iea.org/statistics/?country=WORLD&year=2015&category=Energy consumption&indicator=TFCShareBySector&mode=chart&categoryBrowse=true.

[11] International Energy Agency, “Total Primary Energy Supply by source for World,” *Statistics*, 2017. [Online]. Available: https://www.iea.org/statistics/?country=WORLD&year=2015&category=Key indicators&indicator=TPESbySource&mode=chart&categoryBrowse=false. [Accessed: 19-Jul-2018].

[12] International Energy Agency, “Electricity Generation by fuel,” *Statistics*, 2017. [Online]. Available:

https://www.iea.org/statistics/?country=WORLD&year=2015&category=Electricity&indicator=ElecGenByFuel&mode=chart&categoryBrowse=false. [Accessed: 20-Aug-2018].

[13] S. Ø. Ottesen *et al.*, "INVADE Deliverable 5.3: Simplified Battery Operation and Control Algorithm," INVADE;, 2017.

[14] European Union, "The INVADE project," *INVADE*, 2018. [Online]. Available: https://h2020invade.eu/the-project/.

[15] European Union, "The INVADE Partners," *INVADE*, 2018. .

[16] UPC, "CITCEA About," 2018. [Online]. Available: https://www.citcea.upc.edu/about.html.

[17] Merriam-Webster, "Flexible." 2018.

[18] R. Sims *et al.*, "Integration of Renewable Energy into Present and Future Energy Systems," *IPCC Spec. Rep. Renew. Energy Sources Clim. Chang. Mitig.*, pp. 609–706, 2011.

[19] D. Thomas, O. Deblecker, and C. S. Ioakimidis, "Optimal operation of an energy management system for a grid-connected smart building considering photovoltaics' uncertainty and stochastic electric vehicles' driving schedule," *Appl. Energy*, vol. 210, pp. 1188–1206, 2018.

[20] B. Zhou *et al.*, "Smart home energy management systems : Concept , con fi gurations , and scheduling strategies," vol. 61, pp. 30–40, 2016.

[21] International Energy Agency, "Smart Grid Technology Roadmap," 2011.

[22] Emfores, "Home Energy Management Systems." 2018.

[23] X. Jin, K. Baker, D. Christensen, and S. Isley, "Foresee: A user-centric home energy management system for energy efficiency and demand response," *Appl. Energy*, vol. 205, no. June, pp. 1583–1595, 2017.

[24] R. Hemmati and H. Saboori, "Stochastic optimal battery storage sizing and scheduling in home energy management systems equipped with solar photovoltaic panels," vol. 152, pp. 290–300, 2017.

[25] A. Molderink *et al.*, "Management and Control of Domestic Smart Grid Technology," vol. 1,

no. 2, pp. 109–119, 2010.

[26] R. M. Gonzalez *et al.*, "Optimizing electricity consumption of buildings in a microgrid through demand response," *2017 IEEE Manchester PowerTech, Powertech 2017*, 2017.

[27] A. S. Hassan, L. Cipcigan, and N. Jenkins, "Optimal battery storage operation for PV systems with tariff incentives q," *Appl. Energy*, vol. 203, pp. 422–441, 2017.

[28] M. A. Fotouhi Ghazvini, J. Soares, O. Abrishambaf, R. Castro, and Z. Vale, "Demand response implementation in smart households," *Energy Build.*, vol. 143, pp. 129–148, 2017.

[29] A. Pena-bello, M. Burer, M. K. Patel, and D. Parra, "Optimizing PV and grid charging in combined applications to improve the pro fi tability of residential batteries," *J. Energy Storage*, vol. 13, pp. 58–72, 2017.

[30] A. Mohseni, S. S. Mortazavi, A. Ghasemi, A. Nahavandi, and M. Talaei abdi, "The application of household appliances' flexibility by set of sequential uninterruptible energy phases model in the day-ahead planning of a residential microgrid," *Energy*, vol. 139, pp. 315–328, 2017.

[31] E. Matallanas and D. Masa-bote, "PV self-consumption optimization with storage and Active DSM for the residential sector," vol. 85, no. 2011, pp. 2338–2348, 2014.

[32] M. A. A. Pedrasa, S. Member, T. D. Spooner, and I. F. Macgill, "Coordinated Scheduling of Residential Distributed Energy Resources to Optimize Smart Home Energy Services," vol. 1, no. 2, pp. 134–143, 2010.

[33] IEA, *Harnessing Variable Renewables - a guide to the balancing chalange*. 2011.

[34] H. Klinge Jacobsen and S. T. Schröder, "Curtailment of renewable generation: Economic optimality and incentives," *Energy Policy*, vol. 49, pp. 663–675, 2012.

[35] M. Haller, S. Ludig, and N. Bauer, "Decarbonization scenarios for the EU and MENA power system: Considering spatial distribution and short term dynamics of renewable generation," *Energy Policy*, vol. 47, pp. 282–290, 2012.

[36] International Energy Agency, *The Power to Choose. Demand Response in Liberalised*

Electricity Markets. Paris, 2003.

[37] Merriam-Webster, “Optimization.” 2018.

[38] K. Lange, *Optimization*, Second. 2013.

[39] A. R. Parkinson, R. Balling, and J. D. Hedengren, “Optimization Methods for Engineering Design,” *Brigham Young Univ.*, p. 18, 2013.

[40] S. Ø. Ottesen, *Techno-economic models in Smart Grids: Demand side flexibility optimization for bidding and*, no. March. Halden/Trondheim, 2017.

[41] B. Chachuat, “Mixed-Integer Linear Programming (MILP): Model Formulation.” .

[42] Merriam-Webster, “Risk.” 2018.

[43] A. Tversky and C. R. Fox, “Weighing risk and uncertainty,” *Psychol. Rev.*, 1995.

[44] M. Beaudin, “Residential energy management using a moving window algorithm,” *Innov. Smart Grid …*, pp. 1–8, 2012.

[45] C. Clastres, T. T. Ha Pham, F. Wurtz, and S. Bacha, “Ancillary services and optimal household energy management with photovoltaic production,” *Energy*, vol. 35, no. 1, pp. 55–64, 2009.

[46] D. Livengood and R. Larson, “The Energy Box: Locally Automated Optimal Control of Residential Electricity Usage,” *Serv. Sci.*, vol. 1, no. 1, pp. 1–16, 2009.

[47] S. Koch, D. Meier, M. Zima, M. Wiederkehr, and G. Andersson, “An active coordination approach for thermal household appliances - Local communication and calculation tasks in the household,” *2009 IEEE Bucharest PowerTech Innov. Ideas Towar. Electr. Grid Futur.*, pp. 1–8, 2009.

[48] N. Lu and D. P. Chassin, “A state-queueing model of thermostatically controlled appliances,” *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1666–1673, 2004.

[49] A. H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, “Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid,” *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331,

2010.

[50] J. Lunden, S. Werner, and V. Koivunen, “Distributed demand-side optimization with load uncertainty,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 5229–5232, 2013.

[51] A. J. Conejo, J. M. Morales, and L. Baringo, “Real-time demand response model,” *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 236–242, 2010.

[52] Z. Chen and L. Wu, “Residential appliance DR energy management with electric privacy protection by online stochastic optimization,” *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1861–1869, 2013.

[53] Z. Chen, L. Wu, and Y. Fu, “Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization,” *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1822–1831, 2012.

[54] P. Samadi, H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, “Tackling the load uncertainty challenges for energy consumption scheduling in smart grid,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1007–1016, 2013.

[55] P. Liu and Y. Fu, “Optimal operation of energy-efficiency building: A robust optimization approach,” *IEEE Power Energy Soc. Gen. Meet.*, no. 1, pp. 1–5, 2013.

[56] T. T. Kim and H. V. Poor, “Scheduling Power Consumption With Price Uncertainty,” *IEEE Trans. Smart Grids*, vol. 2, no. 3, pp. 519–527, 2011.

[57] C. A. Correa-florez, A. Gerossier, A. Michiorri, and G. Kariniotakis, “Stochastic operation of home energy management systems including battery cycling,” *Appl. Energy*, vol. 225, no. June, pp. 1205–1218, 2018.

[58] G. Liu, Y. Xu, and K. Tomsovic, “Bidding strategy for microgrid in day-ahead market based on hybrid stochastic/robust optimization,” *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 227–237, 2016.

[59] J. Ombach, “A Short Introduction to Stochastic Optimization,” vol. 23, pp. 9–20, 2014.

[60] J. Watson and D. L. Woodruff, “PYSP User Documentation,” pp. 1–33, 2011.

[61] Ö. Çavu and M. S. Aktürk, “Bounds on risk-averse mixed-integer multi-stage stochastic programming problems with mean-CVaR,” vol. 266, pp. 595–608, 2018.

[62] D. L. Woodruff, J. Watson, and W. E. Hart, “PySP : modeling and solving stochastic programs in Python,” pp. 109–149, 2012.

[63] J. Watson and D. L. Woodruff, “Progressive Hedging for Multi-stage Stochastic Optimization Problems,” pp. 1–15.

[64] S. Jin, S. M. Ryan, J.-P. Watson, and D. L. Woodruff, “Modeling and solving a large-scale generation expansion planning problem under uncertainty,” *Energy Syst.*, vol. 2, no. 3–4, pp. 209–242, 2011.

[65] J. Emmanuel, “Decision Analysis 2: EMV and EVPI- Expected Value and Perfect Information,” Canada, 2015.

[66] L. F. Escudero, A. Garín, M. Merino, and G. Pérez, “The value of the stochastic solution in multistage problems,” *TOP*, 2007.

[67] Statistics Norway, “Electricity Prices,” 2018. [Online]. Available: https://www.ssb.no/en/elkraftpris.

[68] The Data Team, “Python is becoming the world’s most popular coding language,” *The Economist*, 2018. [Online]. Available: https://www.economist.com/graphic-detail/2018/07/26/python-is-becoming-the-worlds-most-popular-coding-language?fsrc=scn/fb/te/bl/ed/pythonisbecomingtheworldsmostpopularcodinglanguagedailychart.

[69] A. Downey, *Think Python*, 2nd ed. Needham: Green Tea Press, 2015.

[70] M. Gardener, *Beggining R. The Statistical Programming Language*. Indianapolis, 2012.

[71] D. Kopf, “The Great R versus Python for data science debate,” *Quartz*, 2017. .

[72] Tidyverse, “Overview,” *dplyr*. .

[73] “Python Data Analysis Library,” *Pandas*, 2018. [Online]. Available: https://pandas.pydata.org/. [Accessed: 13-Aug-2018].

[74] NetworkX Developers, “Software for complex networks,” *NetworkX*, 2018. [Online]. Available: https://networkx.github.io/.

[75] Q. Chen and B. Brunaud, “Pyomo and JuMP – Modeling environments for the 21st Century.” Carnegie Mellon University, 2017.

[76] Pecan Street Inc. Dataport, “Electric Data ID 5357.” 2018.

[77] Pecan Street Inc. Dataport, “Dataport Metadata Dictionary,” 2018.

[78] J. N. Louis and E. Pongrácz, “Life cycle impact assessment of home energy management systems (HEMS) using dynamic emissions factors for electricity in Finland,” *Environ. Impact Assess. Rev.*, 2017.

[79] J. N. Louis, A. Caló, K. Leiviskä, and E. Pongrácz, “Environmental impacts and benefits of smart home automation: Life cycle assessment of Home Energy Management System,” in *IFAC-PapersOnLine*, 2015.

[80] A. Moro and L. Lonza, “Electricity carbon intensity in European Member States: Impacts on GHG emissions of electric vehicles,” *Transportation Research Part D: Transport and Environment*, 2017.

[81] TMB, “TMB en accio, un viatje sostenible.” 2014.

[82] P. Artzener, F. Delbaen, J. M. Eber, and D. Heath, “Coherent Measures of Risk,” *Math. Financ.*, no. 9, pp. 203–228, 1999.

Annexes

Annex 1

Axioms that define Coherent Risk Measures

Definition

1. *Subadditivity:* $R(\xi + \zeta) \leq R(\xi) + R(\zeta)$ for any random variables ξ and ζ;
2. *Positive homogeneity (of degree one):* $R(\lambda\xi) = \lambda R(\xi)$ for all $\lambda \geq 0$;
3. *Monotonicity:* $R(\xi) \leq R(\zeta)$ whenever $\xi \preccurlyeq \zeta$, where $\preccurlyeq$ indicates first-order stochastic dominance, i.e.,$P(\xi \leq t) \geq P(\zeta \leq t), \forall t$;
4. *Translation invariance:* $R(\xi + t) = R(\xi) + t$ for any $t \in \Re$.

[82]

Annex 2

Example on EMV, EVPI and VSS taken from [65] **and based on Birge and Loveaux [3]**

Alternatives	Below Average	Average	Above Average
Wheat	2	2.5	3
Corn	2.4	2.7	3
Sugar Beats	1.7	2.8	3.9
Probabilities	0.3	0.3	0.3

EMV

$$EMV_{Wheat} = \sum_s p_s R_{ds} = 0.3 * 2 + 0.3 * 2.5 + 0.3 * 3 = 2.5$$

$$EMV_{Corn} = \sum_s p_s R_{ds} = 0.3 * 2.4 + 0.3 * 2.7 + 0.3 * 3 = 2.7$$

$$EMV_{Sugar\ Beats} = \sum_s p_s R_{ds} = 0.3 * 1.7 + 0.3 * 2.8 + 0.3 * 3.9 = 2.8$$

$$EMV = \max_d \sum_s p_s R_{ds} = \max_d \left(EMV_{Wheat}, EMV_{Corn}, EMV_{Sugar\ Beats}\right) = 2.8$$

EVPI

$$WS_{BA} = \max_d R_{ds} = 2.4$$

$$WS_{AV} = \max_d R_{ds} = 2.8$$

$$WS_{AA} = \max_d R_{ds} = 3.9$$

$$WS = \sum_s p_s(\max_d R_{ds}) = p_s WS_{BA} + p_s WS_{AV} + p_s WS_{AA} = 0.3 * 2.4 + 0.3 * 2.8 + 0.3 * 3.9 = 3.03$$

$$EVPI = WS - EMV = 3.03 - 2.8 = 0.23$$

VSS

$$VSS = EEV - EMV$$

$$VSS = EEV - 2.8$$

Annex 3

Scripts

Main Program

```
### Main Program###

#Prosumer with battery, PV, load
#solve with PH/EF and json file CONCRETE Model

from pyomo.environ import *
import pandas as pd
import os
import time as ti
import json

from inputs import *

start = ti.time()
case=(filei[13:-5])

os.system("runph -m ReferenceModel.py --default-rho=1.0 --solver=gurobi --solution-
    writer=pyomo.pysp.plugins.jsonsolutionwriter")

with open("ph_solution.json") as f:
              results = json.load(f)

var={}
for s in scen_index:
    var[s]=results["scenario solutions"][s]["variables"]

## CREATING OUTPUTS ##

#Grid operations
buysell_out_index = []
for s in scen_index:
    buysell_out_index.append (s+ '_bought')
    buysell_out_index.append (s+ '_sold')

final_buysell = pd.DataFrame(index=electricity_price.index, columns =buysell_out_index)
for s in scen_index:
    for t in electricity_price.index:
        time=str(t)
        x_s='x_s'+'['+time+']'
        x_b='x_b'+'['+time+']'
        final_buysell[s+ '_bought'][t] = var[s][x_b]['value']
        final_buysell[s+ '_sold'][t] = var[s][x_s]['value']

#Production
gen_out_index = []
for s in scen_index:
    gen_out_index.append (s+ '_prod')

final_gen = pd.DataFrame(index=electricity_price.index, columns =gen_out_index)

for s in scen_index:
    for t in electricity_price.index:
        time = str(t)
        psi_g='psi_g'+'['+time+']'
        final_gen[s+ '_prod'][t] = var[s][psi_g]['value']

#Battery
bat_out_index = []
for bat in batteries.index:
    bat_out_index.append(bat + '_soc')
    bat_out_index.append(bat + '_charge_power')
    bat_out_index.append(bat + '_discharge_power')
```

```

final_batteries=pd.DataFrame(index=electricity_price.index, columns=bat_out_index)

for s in scen_index:
    for bat in batteries.index:
        for t in electricity_price.index:
            time=str(t)
            sigma_soc='sigma_soc'+'['+time+','+bat+']'
            sigma_ch= 'sigma_ch'+'['+time+','+bat+']'
            sigma_dis='sigma_dis'+'['+time+','+bat+']'
            final_batteries[bat + '_soc'][t] = var[s][sigma_soc]['value']
            final_batteries[bat + '_charge_power'][t] = var[s][sigma_ch]['value']
            final_batteries[bat + '_discharge_power'][t] = var[s][sigma_dis]['value’]
#Load
load_out_index = []
for s in scen_index:
    load_out_index.append (s+ '_load')

final_load = pd.DataFrame(index=electricity_price.index, columns =load_out_index)

for s in scen_index:
    for t in electricity_price.index:
        time=str(t)
        omega_l='omega_l'+'['+time+']'
        final_load[s+ '_load'][t] = var[s][omega_l]['value']

#Cost
cost_index= stage_names +['Total cost']
scenario_index=scen_names +['Expected']

cost = pd.DataFrame(index = scenario_index, columns = cost_index)

for s in scen_index:
    for c in stage_index:
        cost[c][s]=results['scenario solutions'][s]['stage costs'][c]
    cost['Total cost'][s]=results['scenario solutions'][s]['cost']

cost['Total cost']['Expected']=results['node solutions']['RootNode']['expected cost']

  #Time
Elapsed = ti.time()- start
t_time= ti.strftime("%H:%M:%S", ti.gmtime(Elapsed))
timey = pd.DataFrame(t_time, index = ["Time"], columns=["[s]"])

writer = pd.ExcelWriter('results2_'+case+'.xlsx')

final_batteries.to_excel(writer,"batteries")
final_gen.to_excel(writer, 'generation')
final_load.to_excel(writer, 'consumption')
final_buysell.to_excel(writer, 'transactions')
timey.to_excel(writer,'time')
cost.to_excel(writer,'cost')
writer.save()
```

Scenario Tree

```
###Scenario Tree Creation###

from pyomo.environ import *
```

```
from pyomo.core import *
import pandas as pd
import networkx as nx

file='stochasticdata.xlsx'
xl = pd.ExcelFile(file)

nodes=xl.parse('Nodes')
stages=xl.parse('Stages')
variables=xl.parse('Variables')

var_index=variables.index
scen_index=nodes.index
stage_index=stages.index
stage_names= stages.index.tolist()
scen_names=nodes.index.tolist()

def pysp_scenario_tree_model_callback():

    from pyomo.pysp.scenariotree.tree_structure_model import ScenarioTreeModelFromNetworkX

    G =nx.from_pandas_edgelist(nodes,'Father','Children',edge_attr=['Probability'],create_using=nx.DiGraph())
    stm = ScenarioTreeModelFromNetworkX(G,edge_probability_attribute="Probability", stage_names=stage_names)

# Declare the variables for each node (or stage)
    for v in var_index:
        stagey=variables.Stage[v]
        dimension=variables.Dimensions[v]
        d=('['+(dimension*('*,'))[:-1]+']')
        stm.StageVariables[stagey].add(v+d)

# Declare the variable or expression object that reports the cost at each time stage
    for s in stage_index:
        stm.StageCost[s] = stages.Cost[s]

    return stm
```

Reference Model

```
### Reference Model ###

from pyomo.environ import *
from pyomo.core import *
from scenariotree import *
from inputs import *


# Creates an instance for each scenario
def pysp_instance_creation_callback(scenario_name, node_names):
    model = ConcreteModel()
    ###SETS###
    # Sets of periods
    model.t = Set(initialize=init_t)
    model.t_restricted = Set(initialize=init_t[1:], within=model.t)

    # Sets of batteries
    model.b = Set(initialize=init_bat)
```

```
    ## Parameters ##

    #Prosumer
    model.p_r_b = init_p_r_b
    model.p_g_b = init_p_g_b
    model.p_tax = init_p_tax
    model.p_vat = init_p_vat
    model.p_r_s = init_p_r_s
    model.p_g_s = init_p_g_s
    model.x_imp = init_x_imp
    model.x_exp = init_x_exp

    #Batteries
    model.p_b_ch = init_p_b_ch
    model.p_b_dis = init_p_b_dis
    model.q_ch = init_q_ch
    model.q_dis = init_q_dis
    model.a_ch = init_a_ch
    model.a_dis = init_a_dis
    model.o_min = init_o_min
    model.o_max = init_o_max
    model.s_ch = init_s_ch
    model.s_dis = init_s_dis
    model.sigma_socin = init_sigma_socin

    #PV
    model.p_g = init_p_g
    model.g_sch = Param(model.t, within=NonNegativeReals, initialize=0.0, mutable=True)

    #Loads
    model.w_l_inf = Param(model.t, within=NonNegativeReals, initialize= 0.0, mutable =True)

    ## Decision Variables ##

    #Prosumer
    model.x_b = Var(model.t, within=NonNegativeReals)
    model.x_s = Var(model.t, within=NonNegativeReals)
    model.delta_b = Var(model.t, within=Binary)
    model.delta_s = Var(model.t, within=Binary)

    #PV
    model.psi_g = Var(model.t, within=NonNegativeReals)

    #loads
    model.omega_l = Var(model.t,within=NonNegativeReals)

    # Batteries
    model.sigma_ch = Var(model.t*model.b, within=NonNegativeReals)
    model.sigma_dis = Var(model.t*model.b, within=NonNegativeReals)
    model.sigma_soc = Var(model.t*model.b, within=NonNegativeReals)


    ##Constraints##

    def balance (model, t):
        return model.psi_g[t]+ sum(model.sigma_dis[t,b] for b in model.b)+model.x_b[t] == model.x_s [t] +model.omega_l[t]+sum (model.sigma_ch[t,b] for b in model.b)
    model.balance = Constraint(model.t, rule =balance)

    #Batteries
```

```
    def soc_evolution(model, t_restr, b):
        return model.sigma_soc[t_restr, b] == model.sigma_soc[t_restr-1, b] + model.a_ch[b] * model.sigma_ch[t_restr,b] - model.sigma_dis[t_restr,b]/model.a_dis[b]
    model.soc_evolution = Constraint(model.t_restricted, model.b, rule=soc_evolution)

    def soc_initial(model,b):
        return model.sigma_soc[0, b] == model.sigma_socin[b] + model.a_ch[b] * model.sigma_ch[0,b] - model.sigma_dis[0,b]/model.a_dis[b]#aqui porque es 1 y no cero, el momento inicial
    model.soc_evolution_init = Constraint(model.b, rule=soc_initial)

    def bat_max_power_in(model, t, b):
        return model.sigma_ch[t,b] <= model.q_ch[b]
    model.bat_max_power_in = Constraint(model.t, model.b, rule=bat_max_power_in)

    def bat_max_power_out(model, t, b):
        return model.sigma_dis[t,b] <= model.q_dis[b]
    model.bat_max_power_out = Constraint(model.t, model.b, rule=bat_max_power_out)

    def bat_max_soc(model, t, b):
        return model.sigma_soc[t,b] <= model.o_max[b]
    model.bat_max_soc = Constraint(model.t, model.b, rule=bat_max_soc)

    def bat_min_soc(model, t, b):
        return model.sigma_soc[t,b] >= model.o_min[b]
    model.bat_min_soc = Constraint(model.t, model.b, rule=bat_min_soc)

    def bat_s_ch(model, t, b):
        return model.sigma_ch[t,b] <= ((-model.q_ch[b])/(1-model.s_ch[b]))*((model.sigma_soc[t,b]/model.o_max[b])-1)
    model.bat_s_ch = Constraint(model.t, model.b, rule=bat_s_ch)

    def bat_s_dis(model, t, b):
        return model.sigma_dis[t,b] <= ((model.q_dis[b]/model.s_dis[b])*((model.sigma_soc[t,b])/(model.o_max[b])))
    model.bat_s_dis= Constraint(model.t, model.b, rule=bat_s_dis)

    #PV
    def PV_max_gen (model, t):
        return model.psi_g [t] <= model.g_sch [t]
    model.PV_max_gen =Constraint (model.t, rule =PV_max_gen)

    #Loads
    def load_inf (model, t):
        return model.omega_l [t]== model.w_l_inf[t]
    model.load_inf =Constraint (model.t, rule =load_inf)

    #Prosumer

    def limit_buy_sell (model, t):
        return (model.delta_b[t] + model.delta_s[t]) <=1
    model.limit_buy_sell =Constraint (model.t, rule =limit_buy_sell)

    def limit_x_buy (model, t):
        return (model.x_b[t])<= model.delta_b[t]*model.x_imp
    model.limit_x_buy =Constraint (model.t, rule =limit_x_buy)

    def limit_x_sell (model, t):
        return (model.x_s[t])<= model.delta_s[t]*model.x_exp
    model.limit_x_sell =Constraint (model.t, rule =limit_x_sell)
```

```python

    ## Objective Function ##

    def first_stage_cost(model):
        battery_charging = sum(model.sigma_ch[i,j]*model.p_b_ch[j] for (i,j) in model.t * model.b)
        battery_discharging = sum(model.sigma_dis[i,j]*model.p_b_dis[j] for (i,j) in model.t * model.b)
        return battery_charging+battery_discharging
    model.FirstStageCost = Expression(rule=first_stage_cost)

    def second_stage_cost(model):
        PV_generation =sum((-model.psi_g[t]+model.g_sch[t])*model.p_g for t in model.t)
        z = sum(((model.p_r_b[i]+model.p_g_b[i]+model.p_tax[i])*model.x_b[i]*model.p_vat-(model.p_r_s[i]+model.p_g_s[i])*model.x_s[i]) for i in model.t)
        return z - PV_generation

    model.SecondStageCost = Expression(rule=second_stage_cost)

    def cost_rule (model):
        return model.FirstStageCost + model.SecondStageCost

    model.objective= Objective(rule=cost_rule, sense=minimize)

    g_sch = {}
    w_l_inf = {}
    for s in scen_names:
        g_sch[s] = init_g_sch[s]
        w_l_inf[s] = init_w_l_inf[s]
    instance = model.clone()
    instance.g_sch.store_values(g_sch[scenario_name])
    instance.w_l_inf.store_values(w_l_inf[scenario_name])

    return instance
```

Inputs

```python
### Inputs File ###

import pandas as pd

filep = 'market.xlsx'
fileb = 'batteries.xlsx'
filei='instance_data_robust.xlsx'
files='stochasticdata.xlsx'

xlp = pd.ExcelFile(filep)
xlb = pd.ExcelFile(fileb)
xli = pd.ExcelFile(filei)
xls = pd.ExcelFile(files)

#Market File
electricity_price=xlp.parse('electricity_price')
prosumer_capacity=xlp.parse('prosumer_capacity')
p_vat=xlp.parse('p_vat')
generator_cost = xlp.parse ('generator_cost')

#Bateries File
batteries=xlb.parse('batteries')

#Stochastic Data File
nodes=xls.parse('Nodes')
```

ETSEIB

```
stages=xls.parse('Stages')
variables=xls.parse('Variables')

var_index=variables.index
scen_index=nodes.index
stage_index=stages.index

stage_names= stages.index.tolist()
scen_names= nodes.index.tolist()

#Instance Data File
forecast_g_r= xli.parse('forecast_g_r')
inflex_load=xli.parse('inflex_load')

#Indexes
init_t = electricity_price.index
init_bat = batteries.index

#Instance parameters

init_g_sch={}
init_w_l_inf={}

for s in scen_names:
    init_g_sch[s] ={(period): forecast_g_r[s][period] for period in electricity_price.index}
    init_w_l_inf[s] ={(period): inflex_load[s][period] for period in electricity_price.index}

#Batteries
init_p_b_ch = batteries['cost_charging'].to_dict()
init_p_b_dis = batteries['cost_discharging'].to_dict()
init_q_ch = batteries['max_charging'].to_dict()
init_q_dis = batteries['max_discharging'].to_dict()
init_a_ch = batteries['efficiency_charge'].to_dict()
init_a_dis = batteries['efficiency_discharge'].to_dict()
init_o_min = batteries['min_storage'].to_dict()
init_o_max = batteries['max_storage'].to_dict()
init_s_ch = batteries['s_ch'].to_dict()
init_s_dis = batteries['s_dis'].to_dict()
init_sigma_socin =batteries ['initial_charge'].to_dict()

#PV
init_p_g = generator_cost ['cost_reduced_generation'][0]

#Prosumer
init_p_r_b = electricity_price['P_retail_buy'].to_dict()
init_p_g_b = electricity_price['P_grid_buy'].to_dict()
init_p_tax = electricity_price['P_tax'].to_dict()
init_p_vat = p_vat['p_vat'][0]
init_p_r_s = electricity_price['P_retail_sell'].to_dict()
init_p_g_s = electricity_price['P_grid_sell'].to_dict()
init_x_imp = prosumer_capacity['imp_cap'][0]
init_x_exp = prosumer_capacity['exp_cap'][0]
```

Creation of Scenarios with R

```
###Creation of Scenarios

```

```r
hourdf<- read.csv("dataport-export-hour.csv",header = TRUE,sep = ";")
library(dplyr)
library(xlsx)
library(ggplot2)

Gen = hourdf %>%
  select(Month,Hour,gen)%>%
  filter(Month=="March")%>%
  select(Hour,gen)
Load = hourdf %>%
  select(Month,Hour,use)%>%
  filter(Month=="March")%>%
  select(Hour,use)

num_scen=10
sense=1

Gen[Gen<0] <-0

Gen_stats = Gen %>%
  group_by(Hour) %>%
  summarize(Average= mean(gen),
            Deviation = sd(gen), Mini= min(gen),
            Maxi = max(gen),n=n())
Gen_stats <-round(Gen_stats, digits=5)
Load_stats = Load %>%
  group_by(Hour) %>%
  summarize(Average= mean(use),
            Deviation = sd(use), Mini= min(use),
            Maxi = max(use),n=n())
Gen_stats <-round(Gen_stats, digits=5)
Load_stats <-round(Load_stats, digits=5)
n = Gen_stats%>%
  summarize(n=n())
h=n[[1]]

Gen_sc <- data.frame(matrix(nrow=h,ncol=num_scen))
g_sc_pr <- data.frame(matrix(nrow=h,ncol=num_scen))
Load_sc <- data.frame(matrix(nrow=h,ncol=num_scen))
l_sc_pr <- data.frame(matrix(nrow=h,ncol=num_scen))
yi <- 1:num_scen
va<-paste0("Scenario",yi)
for (j in 1:h){
  for (i in 1:num_scen){
    mean<- Gen_stats[[j,2]]
    sd <- (sense*Gen_stats[[j,3]])
    if (i<num_scen){
      Gen_sc[j,i]<- qnorm(i/num_scen, mean, sd)
    }else {
      Gen_sc[j,i]<- qnorm(0.99, mean, sd)
    }

    if (i>1){
      g_sc_pr[j,i]<- ((pnorm(Gen_sc[[j,i]],mean,sd)-pnorm(Gen_sc[[j,i-1]],mean,sd)))
    }else {
      g_sc_pr[j,i] <- pnorm(Gen_sc[[j,i]],mean,sd)
    }
  }
  for (i in 1:num_scen){
    mean<- Load_stats[[j,2]]
    sd <- (sense*Load_stats[[j,3]])
    if (i<num_scen){
```

```
        Load_sc[j,i]<- qnorm(i/num_scen, mean, sd)
      }else {
        Load_sc[j,i]<- qnorm(0.99, mean, sd)
      }

      if (i>1){
        l_sc_pr[j,i]<- ((pnorm(Load_sc[[j,i]],mean,sd)-pnorm(Load_sc[[j,i-1]],mean,sd)))
      }else {
        l_sc_pr[j,i] <- pnorm(Load_sc[[j,i]],mean,sd)
      }
    }
}
Gen_sc[Gen_sc<0] <- 0
Gen_sc<-round(Gen_sc, digits=5)
g_sc_pr<-round(g_sc_pr, digits=5)

Load_sc[Load_sc<0] <- 0
Load_sc<-round(Load_sc, digits=5)
l_sc_pr<-round(g_sc_pr, digits=5)

row.names(Gen_sc)<-(as.numeric(row.names(Gen_sc))-1)
row.names(Load_sc)<-(as.numeric(row.names(Load_sc))-1)
colnames(Gen_sc)<-va
colnames(Load_sc)<-va

write.xlsx(Gen_sc,"C:\\Users\\nep19\\Documents\\Tesis\\Código\\Scenarios\\instance_data_s0
   5.xlsx",sheetName="forecast_g_r",col.names = TRUE, row.names=TRUE)
write.xlsx(Load_sc,"C:\\Users\\nep19\\Documents\\Tesis\\Código\\Scenarios\\instance_data_s
   05.xlsx",sheetName="inflex_load",col.names = TRUE, row.names=TRUE, append = TRUE)
```